Institute of Animal Science, University of Hohenheim
Animal Genetics and Breeding (460g)
Prof. Dr. Jörn Bennewitz

# Selection Methods for Local Breeds with Historical Introgression

**DISSERTATION**

submitted in fulfillment of the regulations to acquire the degree "Doktor der Agrarwissenschaften" (Dr.sc.agr./Ph.D. in Agricultural Sciences)

to the

Faculty of Agricultural Sciences

Presented by:

**Yu Wang**

Born in Nei Mongol, China

Stuttgart-Hohenheim, 2018

This thesis was accepted as a doctoral thesis (Dissertation) in fulfillment of the regulations to acquire the doctoral degree “Doktor der Agrarwissenschaften” by the Faculty of Agricultural Sciences at University of Hohenheim on 06. 09. 2018.

| | |
|---|---|
| **Date of the Oral Examination:** | 06. 09. 2018 |
| **Examination Committee:** | |
| Chairperson of the oral examination | Prof. Dr. Markus Rodehutscord |
| Supervisor and Reviewer | Prof. Dr. Jörn Bennewitz |
| Co- Reviewer | Prof. Dr. Theodorus Meuwissen |
| Additional Examiner | Prof. Dr. Mizeck Chagunda |

*Dedicated to my family and friends for their endless*

*support and love*

# TABLE OF CONTENTS

# SUMMARY (ENGLISH)

For the management of local breeds with historical introgression, both genetic gain and the long-term evolution of genetic variability have to be taken into consideration. Traditional optimum contribution selection (traditional OCS) aims at maximizing genetic gain while controlling the rate of inbreeding by optimizing the genetic contribution of each selection candidate to the next generation. It is also a promising approach to maintain genetic diversity since the average kinship of selection candidates is restricted. However, for the breeds with historical introgression, this diversity may be caused by introducing genetic material from other breeds, which can be a risk of the conservation of small local populations. Therefore, the breeding objectives should not only focus on increasing genetic gain but also on maintaining the diversity of native alleles. The main aim of this project was to resolve the existing conflicts in the current breeding program of local breeds with historical introgression.

**Chapter 1** gave a brief introduction and background of the topic and formulated the objective of the thesis. In **chapter 2**, the current inbreeding status of German Angler cattle was evaluated based on both pedigree ($F_{PED}$) and genomic information. The genomic inbreeding coefficients of 182 Angler cattle were estimated via analyzing the genome proportion of run of homozygosity ($F_{ROH}$) and using the genomic relationship matrix ($F_{GRM}$). On average, the inbreeding level of Angler is relatively low compared to the other breeds ($\overline{F_{PED}}: 0.013; \overline{F_{GRM}}: -0.015; \overline{F_{ROH>1Mb}}: 0.031$). Moderate to strong correlations (0.607–0.702) were found between $F_{PED}$ and $F_{ROH}$ based on different length categories of ROH segments. Moreover, it proved that $F_{ROH}$ is a robust estimating method owing to its ability to capture both ancient and recent inbreeding.

Although traditional OCS may achieve higher genetic gain with the restriction of the defined rate of inbreeding, in this case, inbreeding is not the main problem in the current breeding program and the advantage of OCS may be limited since the level of inbreeding may be lower than the threshold. In **chapter 3**, we developed the advanced optimum contribution selection strategy by considering migrant contribution and conditional kinship at native alleles in the OCS procedure. Different scenarios were compared for both functions of production and conservation based on pedigree information. It has been proved that the advanced OCS

approach can effectively maintain the diversity of native alleles and genetic originality while ensuring genetic improvement with appropriate settings of constraint values.

The availability of high-density single-nucleotide polymorphism (SNP) markers provides a solution for achieving accurate estimates of both coancestry and breed composition. In **chapter 4** and **chapter 5**, we evaluated the long-term performance of advanced OCS strategies in both production and conservation function via simulating several subsequent generations based on genomic information. In **chapter 4**, we found that traditional OCS procedure has slight advantages in increasing genetic gain whilst controlling relatedness compared to truncation selection. However, the introgression of foreign genetic material by traditional OCS is not desirable in the local breed conservation. In the long run, constraining migrant contribution and kinship at native alleles in the OCS procedure is a promising approach to increase genetic gain whilst maintaining genetic uniqueness and diversity. **Chapter 5** mimics a conservation program which aims at increasing the value of a breed for conservation by removing exogenous genetic material, maintaining within-breed genetic diversity, and increasing the genetic diversity among breeds. Simply minimizing the exogenous genetic contribution leads to the loss of both within and between population diversity. Moreover, the recovery process ended at a plateau after several generations. The best scenario was able to increase the native contribution from 0.317 to 0.706 before a segment-based kinship level of 0.10 was reached. This scenario maximized the native contribution, constrained the increase in kinship, and the increase in kinship at native alleles. Moreover, it constrained the mean kinship in a multi-breed core set to the current level, which is desirable for the conservation program.

This thesis ends with a general discussion.

# ZUSAMMENFASSUNG

Für das Management lokaler Rassen mit historischer Fremdrasseneinkreuzung müssen sowohl der Zuchtfortschritt als auch die langfristige Entwicklung der genetischen Variabilität berücksichtigt werden. Die traditionelle Selektion anhand optimierter genetischer Beiträge (engl.: traditional optimum contribution selection, kurz: OCS) zielt darauf ab, den Zuchtfortschritt zu maximieren, während gleichzeitig die Inzuchtrate beschränkt wird. Dies geschieht, indem der genetische Beitrag jedes Selektionskandidaten für die nächste Generation optimiert wird. Zudem handelt es sich um einen vielversprechenden Ansatz zur Erhaltung der genetischen Vielfalt, da die durchschnittliche Verwandtschaft der Selektionskandidaten eingeschränkt wird. Für Rassen mit historischer Fremdrasseneinkreuzung kann die Diversität zwischen Rassen jedoch auch durch die Einführung genetischen Materials von anderen Rassen verringert werden, was ein Risiko für die Erhaltung kleiner lokaler Populationen darstellen kann. Daher sollten die Zuchtziele nicht nur auf die Steigerung des Zuchtfortschritts, sondern auch auf die Erhaltung der Vielfalt der nativen Allele ausgerichtet sein. Hauptziel dieses Projektes war es, die bestehenden Konflikte im aktuellen Zuchtprogramm lokaler Rassen mit historischer Fremdrasseneinkreuzung zu lösen.

In **Kapitel 1** wird eine Einführung in die Thematik gegeben und die Zielsetzung der Arbeit formuliert. In **Kapitel 2** wurde der aktuelle Inzuchtstatus des deutschen Rotviehs/Anglers anhand von genomischen und pedigreebasierten Verwandschaftsinformationen ( $F_{PED}$ ) bewertet. Die genomischen Inzuchtkoeffizienten von 182 Angler-Rindern wurden durch die Analyse des Genomanteils der in homozygoten Segmenten liegt (engl.: runs of homozygosity, kurz: ROH) ($F_{ROH}$) und unter Verwendung der genomischen Verwandtschaftsmatrix ($F_{GRM}$) geschätzt. Das Inzuchtniveau der Rasse Angler ist im Durchschnitt verglichen mit anderen Rassen relativ niedrig ($\overline{F_{PED}}: 0.013; \overline{F_{GRM}}: -0.015; \overline{F_{ROH>1Mb}}: 0.031$). Zwischen $F_{PED}$ und $F_{ROH}$ wurden, basierend auf verschiedenen Längenkategorien von ROH-Segmenten, moderate bis starke Korrelationen (0.607-0.702) gefunden. Zudem konnte gezeigt werden, dass $F_{ROH}$ eine robuste Schätzmethode ist, welche sowohl ältere als auch neuer aufgetretene Inzucht erfassen kann.

Obwohl die traditionelle OCS unter Verwendung einer definierten Inzuchtrate einen höheren Zuchtfortschritt erzielt, kann ihr Vorteil im vorliegenden Fall begrenzt sein, da im aktuellen

Zuchtprogramm Inzucht nicht das Hauptproblem darstellt und das Inzuchtniveau unterhalb des Schwellenwertes liegen kann. In **Kapitel 3** haben wir die verbesserte und verallgemeinerte OCS entwickelt, indem wir den Migrationsbeitrag sowie die bedingte Verwandtschaft bei nativen Allelen im OCS-Verfahren berücksichtigten. Auf der Grundlage von Verwandtschaftsinformationen wurden verschiedene Szenarien verglichen. Es wurde nachgewiesen, dass der verbesserte OCS-Ansatz die Vielfalt der nativen Allele und die genetische Eigenständigkeit effektiv aufrechterhalten und gleichzeitig die genetische Verbesserung durch geeignete Einstellungen der Nebenbedingungen sicherstellen kann.

Die Verfügbarkeit von high-density SNP-Markern ermöglicht es, genaue Schätzungen des Verwandtschaftsgrades als auch der Rassenzusammensetzung zu erhalten. In **Kapitel 4 und 5** haben wir die Langzeitauswirkung der verbesserten OCS-Strategien sowohl hinsichtlich des Zuchtfortschritts als auch hinsichtlich der Erhaltung der genetischen Eigenständigkeit untersucht, indem wir mehrere nachfolgende Generationen basierend auf genomischen Informationen simuliert haben. In **Kapitel 4** stellten wir fest, dass das traditionelle OCS-Verfahren gegenüber der *truncation selection* leichte Vorteile im Zuchtfortschritt bei gleichem Inzuchtanstieg hat. Die Einkreuzung von fremdem genetischen Material durch die traditionelle OCS ist jedoch für den Schutz der lokalen Rassen nicht wünschenswert. Auf lange Sicht ist die Begrenzung des Fremdgenanteils und der Verwandtschaft bei nativen Allelen im OCS-Verfahren ein vielversprechender Ansatz, um den Zuchtfortschritt zu steigern und gleichzeitig die genetische Einzigartigkeit und Diversität zu erhalten. **Kapitel 5** untersucht ein Erhaltungszuchtprogramm, das darauf abzielt, den Wert einer Rasse für die Erhaltung zu erhöhen, indem exogenes genetisches Material entfernt wird, die genetische Vielfalt innerhalb der Rasse erhalten bleibt und die genetische Vielfalt zwischen den Rassen erhöht wird. Eine einfache Minimierung des exogenen genetischen Beitrags führt zum Verlust der Vielfalt innerhalb als auch zwischen Rassen. Darüber hinaus stagnierte der Zuchtfortschritt nach mehreren Generationen. Das beste Szenario war in der Lage, den nativen Beitrag von 0,317 auf 0,706 zu erhöhen, bevor ein segmentbasierter Verwandtschaftsgrad von 0,10 erreicht wurde. Dieses Szenario maximierte den nativen Beitrag, beschränkte die Zunahme der Verwandtschaft und die Zunahme der Verwandtschaft bei nativen Allelen. Darüber hinaus beschränkte es die mittlere Verwandtschaft in einer Mehr-Rassen-Population auf das aktuelle Niveau, was für das Erhaltungsprogramm wünschenswert ist.

Diese Thesis endet mit einer allgemeinen und kapitelübergreifenden Diskussion.

# CHAPTER 1

# GENERAL INTRODUCTION

# 1.1 Overview

Local breeds and varieties of domestic animal species have experienced crossing to breeds with high economic value. Such introgression can have both positive and negative consequences for the local livestock populations. The performance of the local breed can be improved via introgressive hybridization. The genetic diversity may increase due to the gene flow from other populations. However, these two benefits are at the expense of losing the genetic originality of the local population, which will lead the local breed to extinction (Todesco *et al.* 2016). Amador *et al.* (2011) confirmed that after several generations without management, even a small introduction of foreign genetic material will rapidly disperse throughout the original population, and this material is difficult to remove. Therefore, current breeding objectives for the local breeds should not only focus on increasing genetic gains, but also maintaining genetic originality and native allele diversity.

For the management of local breeds with historical introgression, three conflicts have to be addressed, *i.e.* the conflict between increasing genetic gain while managing the inbreeding level, the conflict between maintaining genetic diversity while controlling the loss of genetic uniqueness, and the conflict between increasing genetic gain while recovering the original genetic background (Figure 1.1). The traditional approach of optimum contribution selection (traditional OCS) provides a solution to solve the first problem. It attempts to maximize the genetic response while simultaneously restricting the rate of inbreeding by optimizing the genetic contribution of all selection candidates (Meuwissen 1997; Woolliams *et al.* 2015). OCS restricts the average relatedness of the selection candidates, thus the loss of genetic diversity can be controlled in the long run (Eynard *et al.* 2016; Gómez-Romano *et al.* 2016). However, the management using OCS conserves the amount of genetic diversity retained by the population, which also maintains exogenous genes. When the selection is exerted on an economically important trait, the selected superior alleles would mostly come from the migrant breed, which increases the frequencies of exogenous genes. Thus instead of focusing on genetic gain and rate of inbreeding only, the elimination of exogenous genetic material should be considered to help recover the original genetic background.

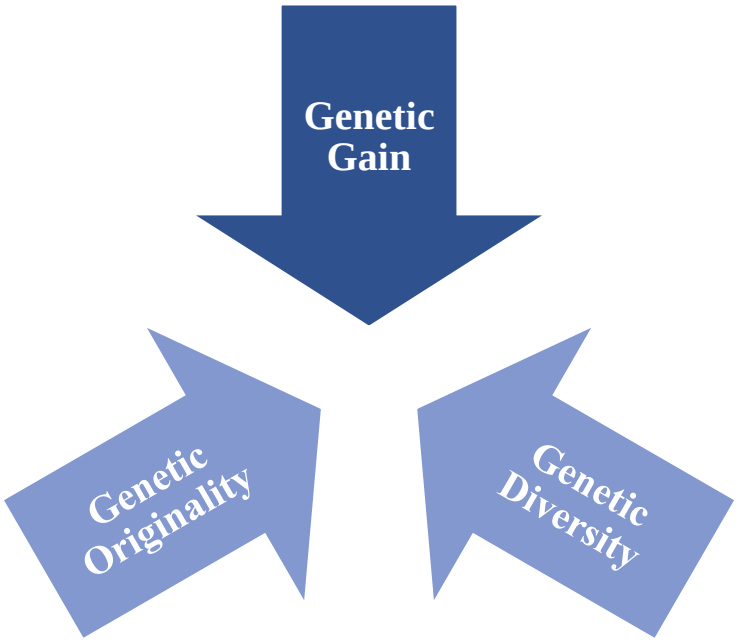



**Figure 1.1** Conflicts in the current breeding programs of local breeds with historical introgression. (Adapted from Wang *et.al.* 2017 EAAP, Tallinn Estonia)

To conserve breeds with historical migrations, Wellmann *et al.* (2012) recommended that approaches should not only constrain migrant contribution (MC), but also aim at increasing the probability that alleles originating from native ancestors are not identical by descent (IBD). Thus the advanced OCS approach, which takes MC and kinship at native alleles as additional constraints, may be a solution to resolve above conflicts to balance both functions of production and conservation.

## 1.2 Study population

In this project, we used the German Angler Rotvieh (shortened as Angler) cattle (see Figure 1.2) as an example in all studies. Angler is a dual purpose cattle breed with an emphasis on dairy production and located in the Northern part of Germany. It is a very old red cattle breed which was first mentioned in the 16$^{th}$ century. Because of the high fat content (4.5-6%) and protein content (3.2-4%) of the milk, the Angler cattle used to be called the 'German Buttercow'. Majority of the bulls are carriers of the kappa casein gene, so the milk has beneficial properties for producing high quality cheese. In addition, this breed was excelled with high adaptability to different environments and farming systems (GGI, German Genetics International GmbH). Currently, there are approximately 13,000 animals registered in the herdbook and genomic selection has already been implemented. Since the late 20$^{th}$ century,

Angler was frequently crossed with other breeds (such as Red Holstein) that the ancient form of this breed is threatened to extinction. This modern type of Angler breed was created through crossbreeding and classified as a separate breed, which was mainly studied in this project.


**Figure 1.2** German Angler cattle (copyright: https://www.rsheg.de/zucht/rassebeschreibung-und-zuchtziele/angler.html)

## 1.3 Objective and outline of the thesis

The overall objective of this thesis was to develop the advanced optimum contribution selection strategies for balancing the need of increasing genetic gain, maintaining genetic diversity and genetic originality to enable more efficient long-term management in both purposes of conservation and production of the breeding program of local livestock breeds.

The thesis is written in chapters and a short description of each chapter is given: **chapter 2** aims at evaluating the current inbreeding status of German Angler cattle based on both pedigree and genotype information. Genomic inbreeding coefficients were evaluated via analyzing the

genome proportion of runs of homozygosity and using the genomic relationship matrix. The linkage disequilibrium status of Angler population was also explored.

In **chapter 3**, the advanced optimum contribution selection strategies were carried out aiming at different objective functions using the example of German Angler cattle and Vorderwald cattle. The analysis was based on pedigree information and single generation. Both migrant contribution and modified kinships that accounts for breed origin of alleles were considered as additional constraints in different OCS scenarios. The traditional pedigree-based kinship was constrained in all optimization scenarios. Not only the targeted function, but also the other parameters were compared among different scenarios.

In **chapter 4,** the long-term performance of different genomic OCS strategies were evaluated using the example of German Angler cattle breed by simulating several subsequent generations. The scenarios were compared not only with respect to the genetic gain but also with respect to parameters measuring genetic diversity and genetic uniqueness. The kinships and migrant contributions, which were evaluated from genomic data, were considered as additional constraints in the OCS scenarios.

**Chapter 5** mimics a conservation program to recovery the original breed after the undesired introgression. Different genomic advanced OCS strategies were designed and multiple subsequent generations were simulated to evaluate the performance of conserving the local German Angler cattle. Two objective functions were considered: to maximize the native contribution and to minimize the kinship among breeds. The performance was evaluated comprehensively, which is not only based on the target objective function, but also the other parameters achieved simultaneously were compared.

This thesis ends with a general discussion.

# Reference

Amador C., Toro M. Á., Fernández J., 2011 Removing exogenous information using pedigree data. Conserv. Genet. 12: 1565–1573.

Eynard S. E., Windig J. J., Hiemstra S. J., Calus M. P. L., 2016 Whole-genome sequence data uncover loss of genetic diversity due to selection. Genet. Sel. Evol. 48: 33.

Gómez-Romano F., Villanueva B., Fernández J., Woolliams J. A., Pong-Wong R., 2016 The use of genomic coancestry matrices in the optimisation of contributions to maintain genetic diversity at specific regions of the genome. Genet. Sel. Evol. 48: 2.

Meuwissen T. H. E., 1997 Maximizing the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75: 2575–2583.

Todesco M., Pascual M. A., Owens G. L., Ostevik K. L., Moyers B. T., *et al.*, 2016 Hybridization and extinction. Evol. Appl. 9: 892–908.

Wellmann R., Hartwig S., Bennewitz J., 2012 Optimum contribution selection for conserved populations with historic migration. Genet. Sel. Evol. 44: 34.

Woolliams J. A., Berg P., Dagnachew B. S., Meuwissen T. H. E., 2015 Genetic contributions and their optimization. J. Anim. Breed. Genet. 132: 89–99.

# CHAPTER 2

# Genome-wide estimates of inbreeding and characteristics of runs of homozygosity in German Angler cattle

Yu Wang[1*], Patrick Stratz[1], Dierck Segelke[2], Robin Wellmann[1], Dirk Hinrichs[3], Jörn Bennewitz[1]

[1]Institute of Animal Science, University of Hohenheim, 70593 Stuttgart, Germany

[2]Vereinigte Informationssysteme Tierhaltung w.V., Heideweg 1, 27283 Verden, Germany

[3]Department of Animal Breeding, University of Kassel, 37213 Witzenhausen, Germany

[*]Corresponding author

In preparation

# Abstract

In this study, we evaluated the genomic inbreeding coefficients based on 182 German Angler cattle via analyzing the genome proportion of runs of homozygosity ($F_{ROH}$) and using the genomic relationship matrix ($F_{GRM}$). Both genomic inbreeding coefficients were compared with the results based on the pedigree information ($F_{PED}$). Moderate to strong correlations (0.607–0.702) were found between $F_{PED}$ and $F_{ROH}$ based on different length categories of ROH segments. $F_{GRM}$ correlates poorly with $F_{PED}$ (0.011) and $F_{ROH}$ (0.096–0.159). The distribution of the ROH segments indicates that $F_{ROH}$ is a robust estimating method owing to its ability to capture both ancient and recent inbreeding. On average, Angler cattle has a relatively low inbreeding level compared to the other breeds ($\overline{F_{PED}}: 0.013; \overline{F_{GRM}}: -0.015; \overline{F_{ROH>1Mb}}: 0.031$), which may be due to the admixture history and the foreign genetic material introduction. This historical introgression process also reflects on its low LD status. The breeding program of Angler cattle should not only focus on achieving higher genetic gain whilst controlling the inbreeding level, but also on recovering its native genetic background to avoid the risk of genetic extinction.

**Keywords**: inbreeding coefficient; runs of homozygosity; linkage disequilibrium; German Angler cattle

# 2.1 Introduction

The inbreeding coefficient is defined as the probability that two randomly chosen alleles at the same locus from an individual are identical by descent (IBD) from a common ancestor (Crow and Kimura 1970). Mating of the related individuals results in inbred offspring, which is inevitable in the populations under selection. The individuals with better performance are usually frequently used in the breeding programs. Subsequently, the rate of inbreeding has increased since the intensive selection process underwent through the populations (Falconer and Mackay 1996). Inbreeding gives rise to allelic fixation and reduces the additive genetic variance. Furthermore, the increased level of inbreeding results in a reduction in the average value of a trait, referred as inbreeding depression, reduces population fitness with increased frequency of recessive deleterious alleles which become homozygous (Howard *et al.* 2017). Thus the management of inbreeding is fundamental to the conservation of genetic resources.

Traditionally, the inbreeding coefficient is estimated based on pedigree records ($F_{PED}$), which is only based on the statistical expectation (VanRaden and Smith 1999). The variation due to Mendelian sampling and linkage during gamete formation cannot be captured. In addition, several parameters may have an influence on the accuracy of the estimation such as the completeness and depth of the pedigree and whether the pedigree contains errors (Cassell *et al.*, 2003). Furthermore, $F_{PED}$ describes the IBD status with respect to the founder generation which assumes individuals are unrelated, which is unrealistic in the real population. The recent evolution of genome sequencing and high-throughput DNA techniques promotes the single nucleotide polymorphism (SNP) arrays an efficient and reliable tool in the era of livestock breeding. It also allows a precise estimate of the realized proportion of the genome that two individuals share (Hill and Weir 2011).

Runs of homozygosity (ROH) are contiguous homozygote segments of the genome where two haplotypes inherited from the parents are identical (Gibson *et al.*, 2006). ROH is proved to be mainly autozygous and the primary cause of ROH is inbreeding (Peripolli *et al.* 2016). Keller *et al.* (2011) indicated that the inbreeding coefficients estimated using ROH is preferable because it could capture both recent and ancient inbreeding. Recent inbreeding from mating closely-related ancestors leads to a high occurrence of long ROH segments. On the other hand, shorter ROH appears when chromosomal segments are broken up by repeated meiosis

(Bjelland *et al.*, 2013; Curik *et al.*, 2014; Ferenčaković *et al.*, 2013). Thus the ROH approach is a reliable tool for estimating the level of inbreeding.

Except for estimating inbreeding coefficient, the use of genotype data also provides a way of evaluating the level of linkage disequilibrium (LD). High linkage disequilibrium can lead to the detection of ROH that are not true IBD. Ferenčaković *et al.* (2013) indicated that there is an abundance of short segments in 50k SNP chip data which is due to LD rather than ancient inbreeding. Thus including such short segments may overestimate the real inbreeding level. It is necessary to check LD status before evaluating the inbreeding level based on ROH.

The aim of the present study was to estimate and compare levels of inbreeding using both genomic and pedigree information of German Angler cattle. Moreover, the status of linkage disequilibrium was also explored.

# 2.2 Material and methods

## 2.2.1 Data

The German Angler Rotvieh cattle (shortened as Angler) is a dual purpose cattle breed mainly located in the northern part of Germany (Bennewitz and Meuwissen 2005). 182 Angler cattle (117 bulls and 65 cows) born between 1986 and 2014 were used in this study. All animals were genotyped with the Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA) containing 45,613 SNP markers. Genotype quality checks were accomplished via using PLINK v1.9 (Chang *et al.* 2015): Markers selected for analysis were required to be located on autosomal chromosomes (866 SNPs were eliminated); only animals with call rates ≥95% and SNP call rate ≥95% were kept in the final dataset (no animal and SNP was eliminated); the SNP markers with minor allele frequency (MAF) <0.01 were discarded (411 SNPs were eliminated); SNP markers with value for the Hardy-Weinberg equilibrium test $<10^{-6}$ were discarded (5 SNPs were eliminated). A total number of 44,331 SNPs were kept as the final genotype dataset for the analysis.

### 2.2.2 Runs of homozygosity detection

ROH segments were detected using PLINK v1.9 (Chang *et al.* 2015), which was designed to find stretches with a specified number of contiguous homozygous SNPs. The following criteria were set to detect the ROH segments: (i) the minimum number of SNPs included in a ROH segment was fixed at 15; (ii) the minimum length of a ROH segment was set at 1Mb; (iii) the maximum distance between adjacent SNPs was 1Mb and the density is at least 1 SNP every 100kb (iv) neither heterozygous nor missing genotypes were allowed in a ROH segment.

### 2.2.3 Pedigree and genomic inbreeding analysis

Three types of inbreeding coefficients ($F_{PED}$, $F_{GRM}$ and $F_{ROH}$) were calculated. The pedigree-based inbreeding coefficient ($F_{PED}$) was estimated using ENDOG software (Gutiérrez and Goyache 2005) with the option that implements the algorithm of Meuwissen & Luo (1992). The genealogy was constructed with all ancestor available for the individuals and included 4,144 animals. The average equivalent complete generations were 7.67 ranging from 4.33 to 9.63.

$F_{GRM}$ was obtained via using the option –ibc from PLINK 1.9 (Chang *et al.* 2015). It is calculated following VanRaden (2008) as below:

$$F_{GRM} = \frac{(x_i - 2\hat{p}_i)^2}{2p_i(1-p_i)} - 1$$

where $p_i$ is the allele frequency at a *i*th locus and $x_i$ is the number of copies of the reference allele for the *i*th SNP. This was equivalent to estimating an individual's relationship to itself (diagonal of the SNP-derived GRM) (Yang *et al.* 2011).

The measure of inbreeding based on ROH ($F_{ROH}$) is defined by McQuillan *et al.* (2008) as the total length of the genome covered by ROH segments divided by the overall length of the genome covered by SNPs as follows:

$$F_{ROH} = \frac{\sum_{j=1}^{nROH} L_{ROH_j}}{L_{total}}$$

where $L_{ROH_j}$ is the length of *j*th ROH and $L_{total}$ is the total length of the genome covered by SNPs, which is 2,521,523 kb in this study. All the results obtained from PLINK were analyzed

using R software version 3.4.1 (R Core Team 2017). For each individual, $F_{ROH}$ was calculated based on ROH of different minimum length: 1, 4, 8 or 16Mb ($F_{ROH>1Mb}$, $F_{ROH>4Mb}$, $F_{ROH>8Mb}$ and $F_{ROH>16Mb}$ respectively), representing up to approximately 50, 13, 6 and 3 generations from common ancestor respectively (Ferenčaković *et al.,* 2013a). In addition, $F_{ROH>1Mb}$ was also calculated for each chromosome.

### 2.2.4 Linkage disequilibrium

Linkage disequilibrium between two SNPs was measured using $r^2$ via PLINK v1.9 (Chang *et al.* 2015). $r^2$ between SNP pairs was defined as $r^2 = \frac{D^2}{p(A)p(1-A)p(B)p(1-B)}$, where $p(A)$, $p(B)$ and $p(AB)$ are the frequencies of allele A, allele B and haplotype AB respectively and $D^2 = (p_{AB} - p_A p_B)^2$ is the squared difference observed and expected haplotype frequency. $r^2$ varies between 0 and 1, where 0 means the SNPs are completely uncorrelated while 1 means the SNPs are completely correlated. For each chromosome, pairwise $r^2$ was calculated for SNPs between 0 and 5Mb apart. The $r^2$ was grouped into bins based on the distance between SNPs from physical map.

## 2.3 Results

### 2.3.1 Runs of homozygosity detection and distribution

The basic information of each chromosome was shown in Table 2.S1 [See Additional file 1 Table 2.S1]. The average distance between SNPs was ~56.70 kb (± 65.06) and the maximum distance between SNPs was 7.12 Mb which was located on chromosome 6.

The number of ROH per animal varied among different ROH length categories, which can be seen in Table 2.1. In the study population, 1854 ROH segments were detected in total. On average 10.19 (±5.51) ROH segments were found per animal (from 0 to 33 in individual samples). The mean sum of all ROH lengths per animal was estimated at 77.88 (±52.81) Mb, with a minimum value of 0 and a maximal value of 304.27 Mb. The number of ROH segments detected in each chromosome is shown in Figure 2.1. The highest numbers of ROH segments were observed on chromosome 1, 3 and 13, in which the number of ROH was 118, 115 and 92 respectively. And the least numbers of ROH were observed on chromosome 23, 28 and 27,

which the number of ROH segments was 18, 25 and 29 respectively. Chromosome 3 has the highest number of both short and long ROH segments across the genome that 47 ROH segments with the length of 1-4 Mb and 15 ROH segments with the length of >16Mb were detected. A detailed description can be found in Table 2.S2 [See Additional file 2 Table 2.S2].

**Table 2.1** Descriptive statistics of the number and length of ROH segments in selected ROH categories.

| | Statistics | ROH length category (Mb) | | | |
|---|---|---|---|---|---|
| | | 1+ | 4+ | 8+ | 16+ |
| Total number of ROH segments ($n_{ROH}$) | n | 1,854 | 1,193 | 536 | 170 |
| Number of ROH segments per animal ($n_{ROH}$) | Mean | 10.19 | 6.56 | 2.95 | 0.93 |
| | SD | 5.51 | 4.23 | 2.46 | 1.22 |
| | Min | 0 | 0 | 0 | 0 |
| | Max | 33 | 20 | 13 | 6 |
| Length of ROH segments per animal (Mb) | Mean | 77.88 | 66.91 | 46.63 | 24.49 |
| | SD | 52.81 | 50.39 | 43.55 | 33.83 |
| | Min | 0 | 0 | 0 | 0 |
| | Max | 304.27 | 269.06 | 219.87 | 193.60 |

The number of ROH segments and the length covered by ROH segments per individual is considerably different among animals, which is shown in Figure 2.2. Both short and long ROH segments were distributed in the genome. Same number of ROH segments detected in different animals may accumulated to different length. For example for animals which have 5 ROH segments detected in the genome, the length of the genome covered by ROH segments is ranging from 19.15 to 146.84 Mb. In addition, there is one extreme individual which has 33 ROH segments detected and 304.27 Mb was covered by ROH segments across the genome.

## 2.3.2 Inbreeding coefficients

The basic statistical description of the inbreeding coefficients estimated using different approaches is shown in Table 2.2. On average, $F_{PED}$ and $F_{ROH}$ are higher than $F_{GRM}$. The $F_{ROH}$ values ranged from 0 to 0.121, with a mean of 0.031 (±0.021) when the minimum length of ROH was set as 1Mb. The $F_{PED}$ values were lower, and ranged from 0 to 0.076, with a mean of 0.013 (±0.013). On average the $F_{GRM}$ value was the lowest with a mean of -0.015 (±0.043). The chromosomal-wide level of $F_{ROH>1Mb}$ is shown in Figure 2.S1 [See Additional file 3 Figure 2.S1]. The average inbreeding level of each chromosome ranged from 0.012 (Chromosome 23) to 0.047 (Chromosome 29).

**Table 2.2** Level of inbreeding estimated based on different methods.

| | $F_{PED}$[1] | $F_{GRM}$ | $F_{ROH>1Mb}$ | $F_{ROH>4Mb}$ | $F_{ROH>8Mb}$ | $F_{ROH>16Mb}$ |
|---|---|---|---|---|---|---|
| Mean | 0.013 | -0.015 | 0.031 | 0.027 | 0.019 | 0.010 |
| SD | 0.013 | 0.043 | 0.021 | 0.020 | 0.017 | 0.013 |
| Min | 0 | -0.118 | 0 | 0 | 0 | 0 |
| Max | 0.076 | 0.183 | 0.121 | 0.107 | 0.087 | 0.077 |

[1]$F_{PED}$ is the inbreeding coefficient estimated from pedigree data. $F_{GRM}$ is GRM-based inbreeding coefficient. $F_{ROH}$ is the inbreeding coefficient based on ROH at different length setting

The Pearson's correlations among three categories of inbreeding coefficients are shown in Table 2.3. $F_{GRM}$ has the lowest correlation with both $F_{PED}$ and $F_{ROH}$ in all length categories. $F_{PED}$ has a moderate to high correlation with $F_{ROH}$ in all categories ranging from 0.607 to 0.702. The correlation decreased when the minimum length of ROH segments increase. The correlations between $F_{ROH}$ in different ROH length categories were high among each other, ranging from 0.825 to 0.992.

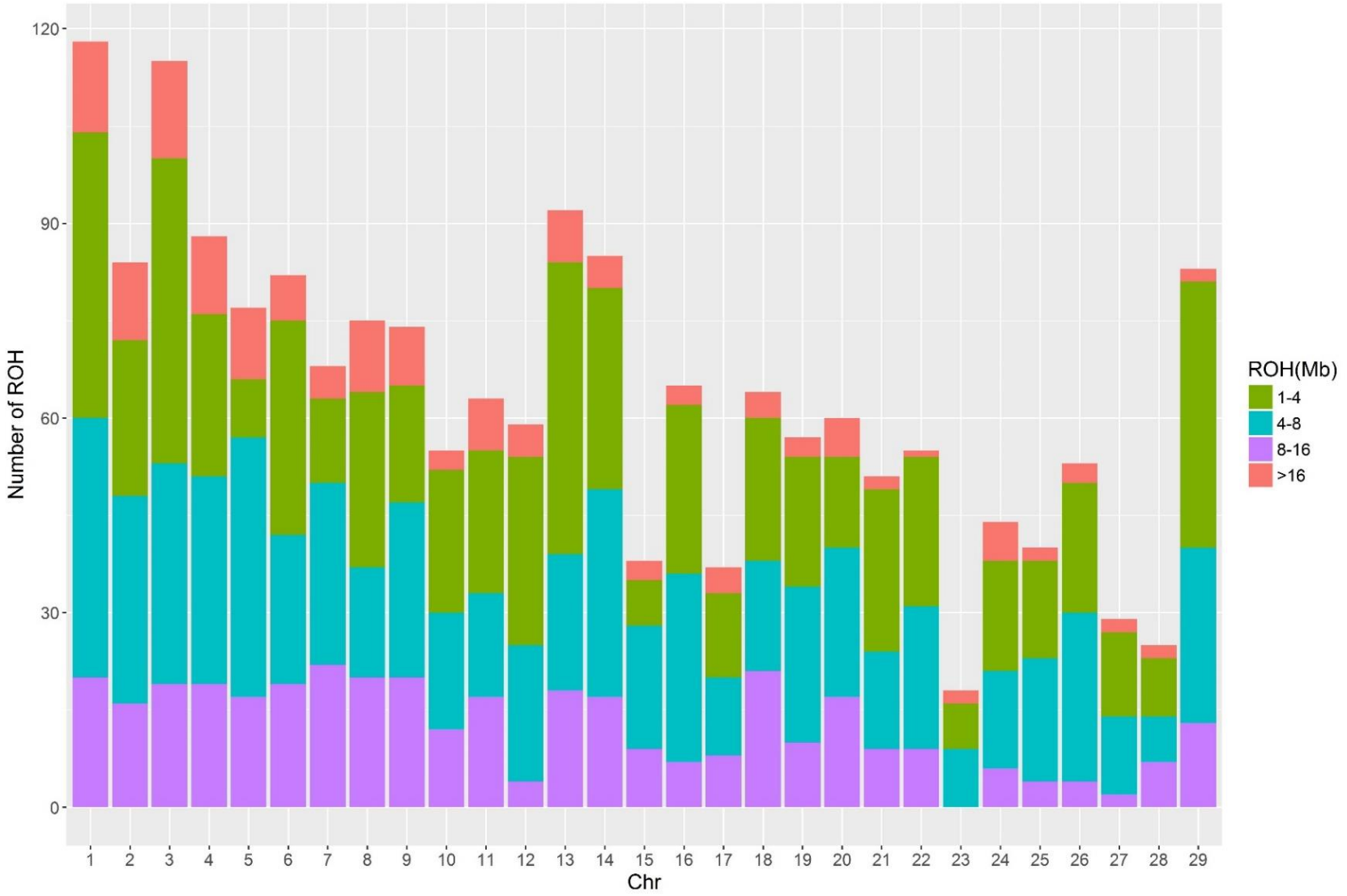



**Figure 2.1** The number of runs of homozygosity (ROH) in different categories detected in each chromosome.

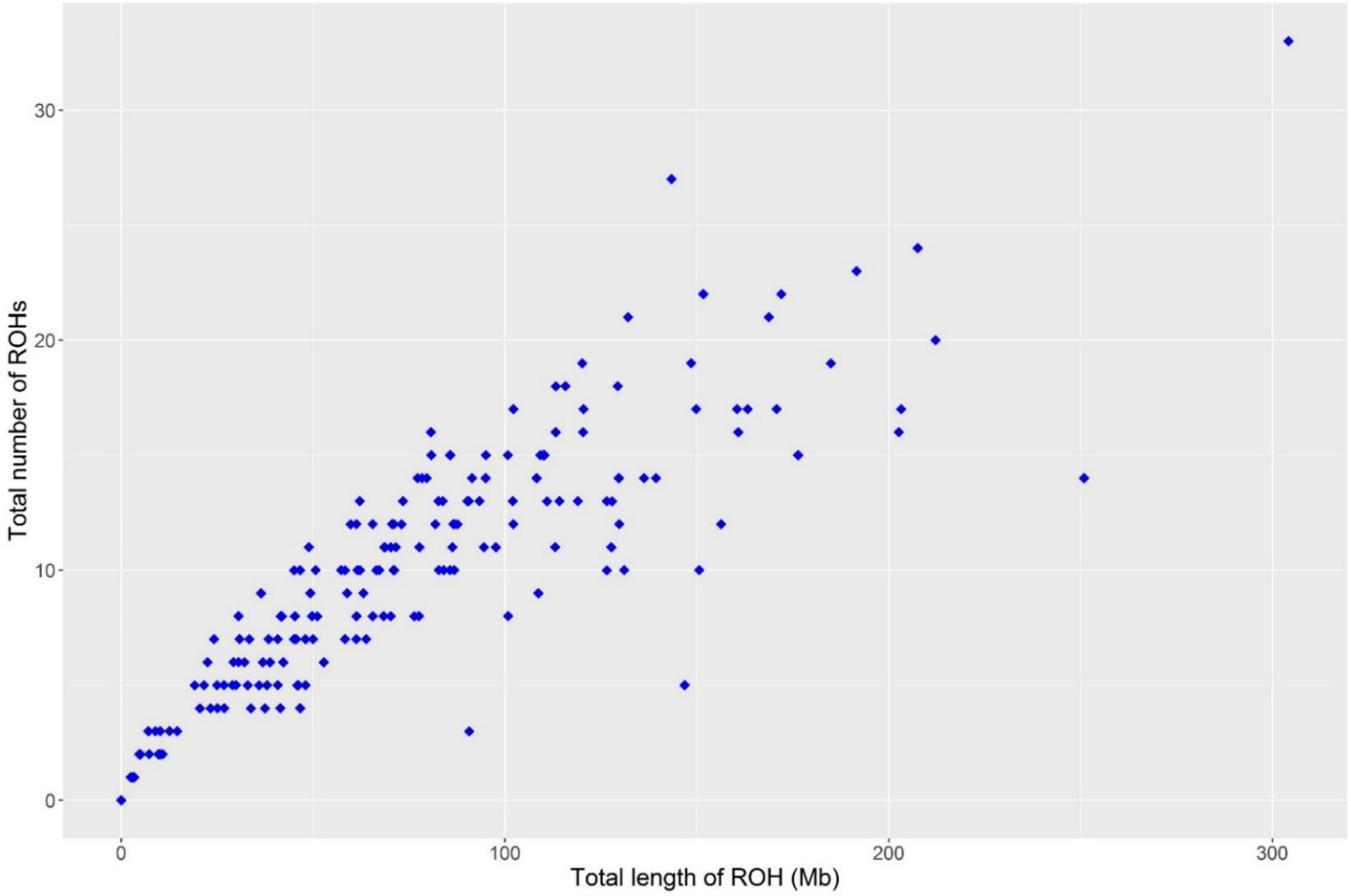



**Figure 2.2** The total number of runs of homozygosity longer than 1MB and the total length of the genome (Mb) covered by ROH segments per individual

**Table 2.3** Pearson's correlation coefficients between different estimates of inbreeding coefficient

| **Correlation** | **$F_{PED}$** | **$F_{GRM}$** | **$F_{ROH>1Mb}$** | **$F_{ROH>4Mb}$** | **$F_{ROH>8Mb}$** | **$F_{ROH>16Mb}$** |
|---|---|---|---|---|---|---|
| $F_{PED}$[1] | 1 | 0.011 | 0.702** | 0.692** | 0.697** | 0.607** |
| $F_{GRM}$ | | 1 | 0.159* | 0.154* | 0.115 | 0.096 |
| $F_{ROH>1Mb}$ | | | 1 | 0.992** | 0.944** | 0.825** |
| $F_{ROH>4Mb}$ | | | | 1 | 0.956** | 0.848** |
| $F_{ROH>8Mb}$ | | | | | 1 | 0.895** |
| $F_{ROH>16Mb}$ | | | | | | 1 |

*: significantly different from 0 at $p < 0.05$; **: significantly different from 0 at $p < 0.01$.
[1]$F_{PED}$ is the inbreeding coefficient estimated from pedigree data. $F_{GRM}$ is GRM-based inbreeding coefficient. $F_{ROH}$ is the inbreeding coefficient based on ROH at different length setting

Figure 2.3 (Above) shows the percentage of the animals in the population which has the SNP in the ROH across the genome. The high proportion of animals has the SNP in the ROH is concentrated in BTA 26. Figure 2.3 (Below) is an exaggerated figure of BTA 26 shows the overlay regions of ROH segments identified in Angler cattle and each row represents one animal. There are 10 consecutive SNPs in BTA 26 that 25 animals have ROH segments on these SNPs located between 21409429 bp to 22118554 bp. But still, the proportion is rather low (13.74%). Moreover, BTA 14 and BTA 18 also have high numbers of animal have ROH segments shared on the SNP. The values in detail can be found in Table 2.S3 [see Additional file 4 Table 2.S3].

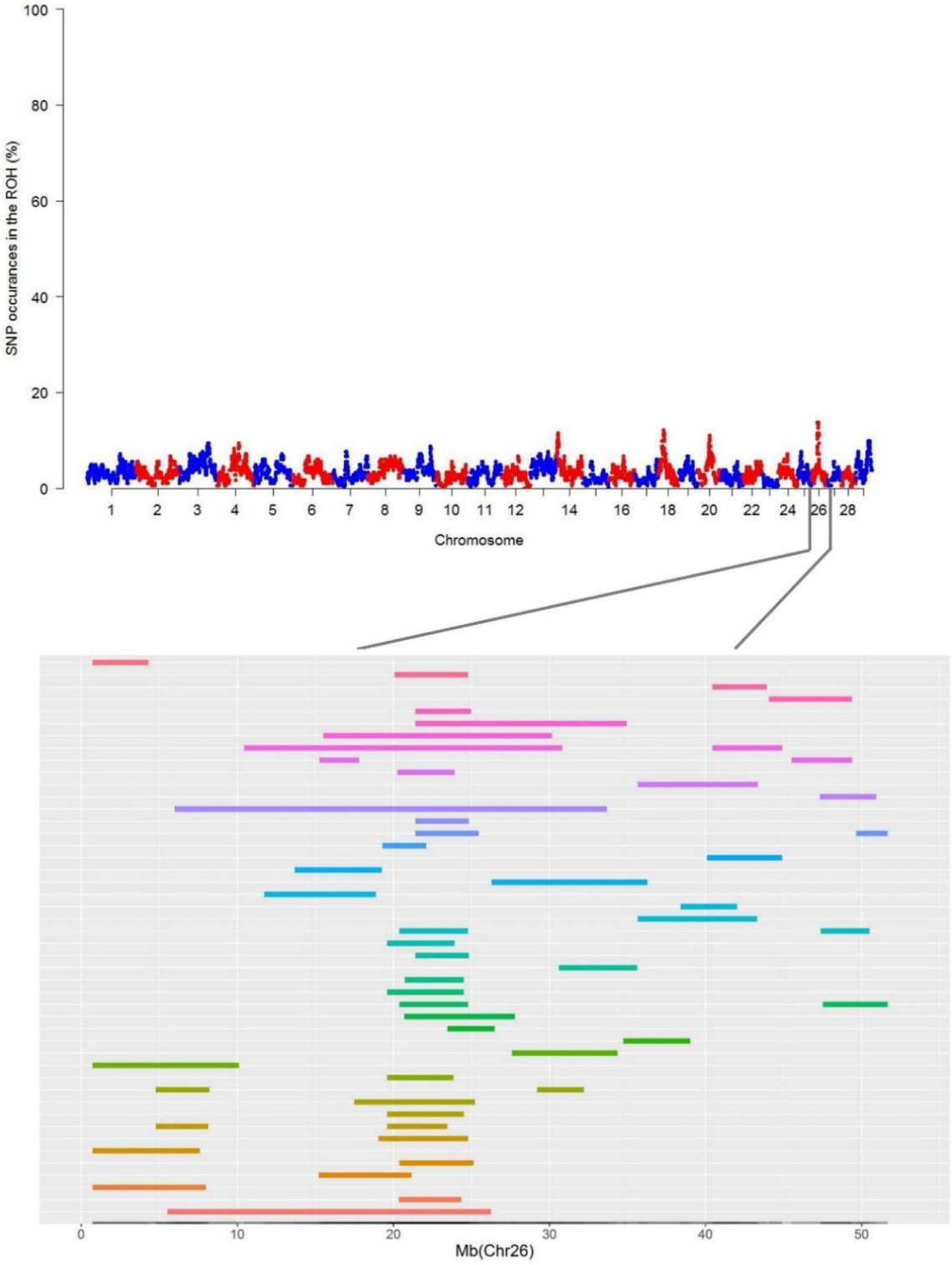



**Figure 2.3** (Above) Proportion of animals with this SNP within an ROH segment. (Below) Overlay region of ROH segments identified on chromosome 26 in Angler cattle. Each row represents one animal.

### 2.3.3 Linkage disequilibrium analysis

All the possible SNP pairs with distances <5Mb on the same chromosome produced 397,616 pairwise LD values across the genome. The average $r^2$ of the adjacent SNP pairs of each chromosome ranged from 0.132 (BTA27) to 0.197 (BTA 13). On average $r^2$ of the adjacent SNP pairs across the genome was 0.162 (±0.221) [See Additional file 1 Table 2.S1]. The mean $r^2$ values and the proportion of SNP pairs that are greater than 0.3 were shown in Table 2.4. A mean value of 0.191 (±0.248) was observed in the pairwise distance of <30kb. It dropped to 0.152 when the pairwise distance increased to 70kb and continued to decrease with the increase of the SNP pairwise distances. Similarly the number of SNP pairs with $r^2 > 0.3$ also decreased with the increase of the SNP pairwise distances.

**Table 2.4** Linkage disequilibrium (LD) over different distances.

| Distance range (kb) | Number of SNP pairs | Number of SNP pairs $r^2$ >0.3 | | Average $r^2$(SD) |
|---|---|---|---|---|
| | | Number | % | |
| 0-30 | 14,540 | 3,198 | 21.994 | 0.191(0.248) |
| 30-70 | 34,543 | 5,749 | 16.643 | 0.152(0.210) |
| 70-100 | 25,444 | 2,752 | 10.816 | 0.114(0.170) |
| 100-200 | 83,235 | 5,127 | 6.160 | 0.084(0.134) |
| 200-500 | 191,343 | 5,504 | 2.877 | 0.061(0.100) |
| 500-1000 | 45,590 | 814 | 1.785 | 0.052(0.086) |
| 1000-5000 | 2,921 | 46 | 1.575 | 0.049(0.076) |

## 2.4 Discussion

In this study, we calculated the inbreeding coefficients of German Angler cattle based on both pedigree and genomic information. The distribution of detected ROH segments was characterized and the linkage disequilibrium status was also evaluated. ROH approach has been proved to be a reliable tool for capturing both ancient and recent inbreeding. Besides it showed that Angler cattle has a relatively low inbreeding level and low LD status.

Several parameters may have an influence on the accuracy of estimating inbreeding coefficients via runs of homozygosity: the minimum length of a ROH segment; the minimum number of SNPs within a ROH segment; the minimum marker density of a ROH segment; the maximum distance between two adjacent SNPs; the maximum number of missing genotypes and the maximum number of heterozygous SNPs admitted within a ROH segment (Peripolli *et al.* 2016). Due to the lack of consensus in the detecting criteria, it is hard to compare the results of different studies based on different species. Ramilo and Fernández (2016) suggested that for 50k data in cattle, a ROH segment should fulfill the following criteria to ensure a high correlation between the pedigree-based inbreeding and the proportion of homozygous SNPs: a length between 0.2-4 Mb, with a minimum number of SNPs less than 80, with a minimum density of 1 SNP every 90KB and a maximum distance between two adjacent SNPs higher than 0.4 Mb. Besides, the maximum number of missing genotypes and the maximum number of heterozygous SNPs allowed in the ROH have no effect on the identification of ROH.

Ferenčaković *et al.* (2013b) mentioned that the 50k panel revealed an abundance of small segments caused by LD. Thus the number of segments of 1-4Mb long accounting for inbreeding may be overestimated. In our study, although paired t-test showed that on average the $F_{ROH>1Mb}$ is significantly higher than the $F_{ROH>4Mb}$ (results not shown), the correlation between $F_{ROH>1Mb}$ and $F_{PED}$ is also higher than $F_{ROH>4Mb}$ and $F_{PED}$. It is possible that the abundance of small segments was because of the ancient inbreeding instead of LD since low LD status was detected in this study. Thus while using 50K data in cattle, it is an arbitrary decision of setting the minimum length of ROH segments 4Mb in estimating inbreeding level ignoring the LD status and heterozygous level of the population. As a consequence, it is recommended to evaluate the LD status prior to determining the minimum length of a ROH segment. For breeds with relatively low LD value and higher genetic diversity, short ROH segments may also be included accounting for ancient inbreeding level.

The proportion of the genome present in ROH segments offers a good indication of inbreeding levels and the analysis based on ROH segment length can distinguish the relative amounts of autozygosity due to recent or remote ancestors (Ferenčaković *et al.*, 2013a). Inbreeding coefficients estimated from pedigree was lower than the estimates of $F_{ROH>1Mb}$. It may be explained by $F_{ROH>1Mb}$ captures both recent and ancient relatedness, while $F_{PED}$ is only extended to a limited number of generations based on the recorded pedigree. If only recent inbreeding is considered ($F_{ROH>16Mb}$), on average the results of $F_{ROH}$ (0.010±0.013) and $F_{PED}$ (0.013±0.013) are rather similar.

Several studies have estimated the inbreeding coefficients using runs of homozygosity. A high correlation between pedigree-based inbreeding and the ROH-based inbreeding is the desired feature of evaluating inbreeding via ROH (Nothnagel *et al.*, 2010). Although each study has its own criteria for detecting ROH, strong to moderate correlations between $F_{PED}$ and $F_{ROH}$ in different cattle breeds were reported by several studies: 0.82 for Holstein cattle and 0.54 for Danish Red cattle (Zhang *et al.*, 2015), 0.70 for Jersey cattle (Kim *et al.*, 2015); 0.659 for Red Holstein and 0.609 for Holstein Frisian (Signer-Hasler *et al.* 2017); 0.64 to 0.72 in Tyrol cattle, 0.52 to 0.70 in Fleckvieh cattle, 0.50 - 0.67 in Brown Swiss cattle, 0.50-0.62 in Norwegian Red cattle (Ferenčaković *et al.*, 2013a). Similarly in our study, moderate to high correlations were found between $F_{PED}$ and $F_{ROH}$ based on different ROH length categories (0.607–0.702). $F_{GRM}$ has a poor correlation with both $F_{PED}$ (0.011) and $F_{ROH}$ (0.096–0.159), which may be due to the bias caused by minor allele frequencies (Zhang *et al.* 2015). Because of the estimate of $F_{GRM}$ is sensitive to allele frequencies and sampling, and the reliability of $F_{PED}$ relies on the pedigree quality and completeness, compare to the other two estimating methods, the advantage of $F_{ROH}$ is obvious that it directly reflects the real level of homozygosity.

In this study, we applied three approaches to estimate the inbreeding coefficients. Based on the pedigree information, the inbreeding coefficient of Angler population was 0.013, which was slightly lower than the previous study (0.02) (Hinrichs and Thaller 2013). Based on the different setting of length categories of ROH segments, the highest average of German Angler cattle $F_{ROH}$ was only 0.031 (±0.021). This low inbreeding level was in accordance with the former study that German Angler cattle has an effective population size of 237.8 (Simianer 2005). In all categories of different ROH length, the inbreeding coefficients of German Angler cattle are all lower than the results from former studies of other Red cattle breeds such as the Danish Red (Zhang *et al.* 2015), Norwegian Red (Hamzic, 2012; Hillestad *et al.*, 2014), Polish

Red (Szmatoła *et al.* 2016) etc. This is likely due to elevated heterogeneity from historic admixture and gene flow through the import of sires from other breeds. German Angler cattle is known to be an admixed breed originated from old German Rotvieh cattle breed with an introgression history with Red Holstein cattle since the 1980s (GGI, German Genetics International GmbH).

The historical admixture process also reflects in the low level of LD. Although in general the average $r^2$ decreased with increasing distance between pairs of markers, for SNP pairs which have short distances, the average $r^2$ of Angler population is still lower than the other cattle breeds (Rius-Vilarrasa *et al.,* 2011; Sodeland *et al.,* 2011). In populations with high LD and recent inbreeding. the 54K SNP panel may provide a good estimate of inbreeding, but in populations with low LD and ancient inbreeding, higher marker density would be needed to identify short ROH segments that are IBD (Marras *et al.* 2015).

# 2.5 Conclusion

In this study, we estimated inbreeding coefficients of German Angler cattle based on both pedigree and genomic information. The length and distribution of ROH segments and LD status of Angler cattle were also explored. There is a moderate to high correlation between inbreeding coefficients estimated from pedigree and ROH which indicates that ROH is a robust method of capturing both ancient and recent inbreeding. The observed inbreeding status was higher based on ROH than was expected from pedigree information. Both estimates from pedigree and genomic information indicated German Angler cattle is a breed with a relatively low inbreeding level and low LD status, which may be caused by previous admixture process.

## Acknowledgment

The study was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG).

## Competing interests

The authors declare that they have no competing interests.

# References

Bennewitz J., Meuwissen T. H. E., 2005 Estimation of extinction probabilities of five German cattle breeds by population viability analysis. J. Dairy Sci. 88: 2949–2961.

Bjelland D. W., Weigel K. A., Vukasinovic N., Nkrumah J. D., 2013 Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. J. Dairy Sci. 96: 4697–4706.

Cassell B. G., Adamec V., Pearson R. E., 2003 Effect of incomplete pedigrees on estimates of inbreeding and inbreeding depression for days to first service and summit milk yield in Holsteins and Jerseys. J. Dairy Sci. 86: 2967–2976.

Chang C. C., Chow C. C., Tellier L. C. A. M., Vattikuti S., Purcell S. M., *et al.*, 2015 Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4: 7.

Crow J. F., Kimura M., 1970 *An Introduction to Population Genetics Theory*. New York, Evanston and London: Harper & Row, Publishers.

Curik I., Ferenčaković M., Sölkner J., 2014 Inbreeding and runs of homozygosity: A possible solution to an old problem. Livest. Sci. 166: 26–34.

Falconer D. S., Mackay T. F. C., 1996 *Introduction to quantitative genetics*. 4th ed Harlow: Pearson Ediucation Limited.

Ferenčaković M., Hamzić E., Gredler B., Solberg T. R., Klemetsdal G., *et al.*, 2013a Estimates of autozygosity derived from runs of homozygosity: Empirical evidence from selected cattle populations. J. Anim. Breed. Genet. 130: 286–293.

Ferenčaković M., Sölkner J., Curik I., 2013b Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. Genet. Sel. Evol. 45: 42.

Gibson J., Morton N. E., Collins A. R., 2006 Extended tracts of homozygosity in outbred human populations. Hum. Mol. Genet. 15: 789–795.

Gutiérrez J. P., Goyache F., 2005 A note on ENDOG: A computer program for analysing pedigree information. J. Anim. Breed. Genet. 122: 172–176.

Hamzic E., 2012 Levels of Inbreeding Derived from Runs of Homozygosity: A Comparison of Austrian and Norwegian Cattle Breeds. Master Thesis, BOKU Vienna

Hill W. G., Weir B. S., 2011 Variation in actual relationship as a consequence of Mendelian sampling and linkage. Genet. Res. (Camb). 93: 47–64.

Hillestad B., Woolliams J. A., Meuwissen T. H. E., Våge D. I., Klemetsdal G., 2014 Estimating Rate of Inbreeding and Effective Population Size using Genomic Data in Norwegian Red Cattle. In: *Proceedings of the 10th world congress on genetics applied to livestock production:* 17–22 August 2014; Vancouver.

Hinrichs D., Thaller G., 2013 Inbreeding and impact of foreign dairy cattle breeds in the German Angeln dairy cattle population. In: *Proceedings of the 64th annual meeting of the European Association for Animal Production (EAAP):*26-30 August 2013; Nantes

Howard J. T., Pryce J. E., Baes C., Maltecca C., 2017 Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. J. Dairy Sci. 100: 1–16.

Keller M. C., Visscher P. M., Goddard M. E., 2011 Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics 189: 237–249.

Kim E.-S., Sonstegard T. S., Tassell C. P. Van, Wiggans G. R., Rothschild M. F., 2015 The relationship between runs of homozygosity and inbreeding in Jersey cattle under selection. PLoS One 10: 1–17.

Marras G., Gaspa G., Sorbolini S., Dimauro C., Ajmone-Marsan P., *et al.*, 2015 Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim. Genet. 46: 110–121.

McQuillan R., Leutenegger A.-L., Abdel-Rahman R., Franklin C. S., Pericic M., *et al.*, 2008 Runs of homozygosity in European populations. Am. J. Hum. Genet. 83: 359–372.

Meuwissen T. H. E., Luo Z., 1992 Computing inbreeding coefficients in large populations. Genet. Sel. Evol. 24: 305–313.

Nothnagel M., Lu T. T., Kayser M., Krawczak M., 2010 Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. Hum. Mol. Genet. 19: 2927–2935.

Peripolli E., Mudadu M. de, Silva M. V. B., Lima A. L. F., Irgang R., *et al.*, 2016 Runs of homozygosity: current knowledge and applications in livestock. Anim. Genet. 48: 255–271.

R Core Team, 2017 R: A language and environment for statistical computing.

Ramilo S. T. R., Fernández J., 2016 What do we mean by runs of homozygosity? Assessing effect of parameters involved in their detection. In: *Proceedings of the 67th annual meeting of the European Association for Animal Production (EAAP):* 29 Aug – 2 Sep 2016; Belfast

Rius-Vilarrasa E., Iso-Touru T., Strandén I., Schulman N., Guldbrandtsen B., *et al.*, 2011 Characterization of linkage disequilibrium in a Danish, Swedish and Finnish Red Breed cattle population. In: *Proceedings of the 62nd annual meeting of the European Association for Animal Production (EAAP):* 29 Aug – 2 Sep 2011; Stavanger

Signer-Hasler H., Burren A., Neuditschko M., Frischknecht M., Garrick D. J., *et al.*, 2017 Population structure and genomic inbreeding in nine Swiss dairy cattle populations. Genet. Sel. Evol. 49: 83.

Simianer H., 2005 Using expected allele number as objective function to design between and within breed conservation of farm animal biodiversity. J. Anim. Breed. Genet. 122: 177–187.

Sodeland M., Kent M., Hayes B. J., Grove H., Lien S., 2011 Recent and historical recombination in the admixed Norwegian Red cattle breed. BMC Genomics 12: 33.

Szmatoła T., Gurgul A., Ropka-Molik K., Jasielczuk I., Ząbek T., *et al.*, 2016 Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland. Livest. Sci. 188: 72–80.

VanRaden P. M., Smith L. A., 1999 Selection and Mating Considering Expected Inbreeding of Future Progeny. J. Dairy Sci. 82: 2771–2778.

VanRaden P. M., 2008 Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414–23.

Yang J., Lee S. H., Goddard M. E., Visscher P. M., 2011 GCTA: A tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88: 76–82.

Zhang Q., Calus M. P. L., Guldbrandtsen B., Lund M. S., Sahana G., 2015 Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. BMC Genet. 16: 88.

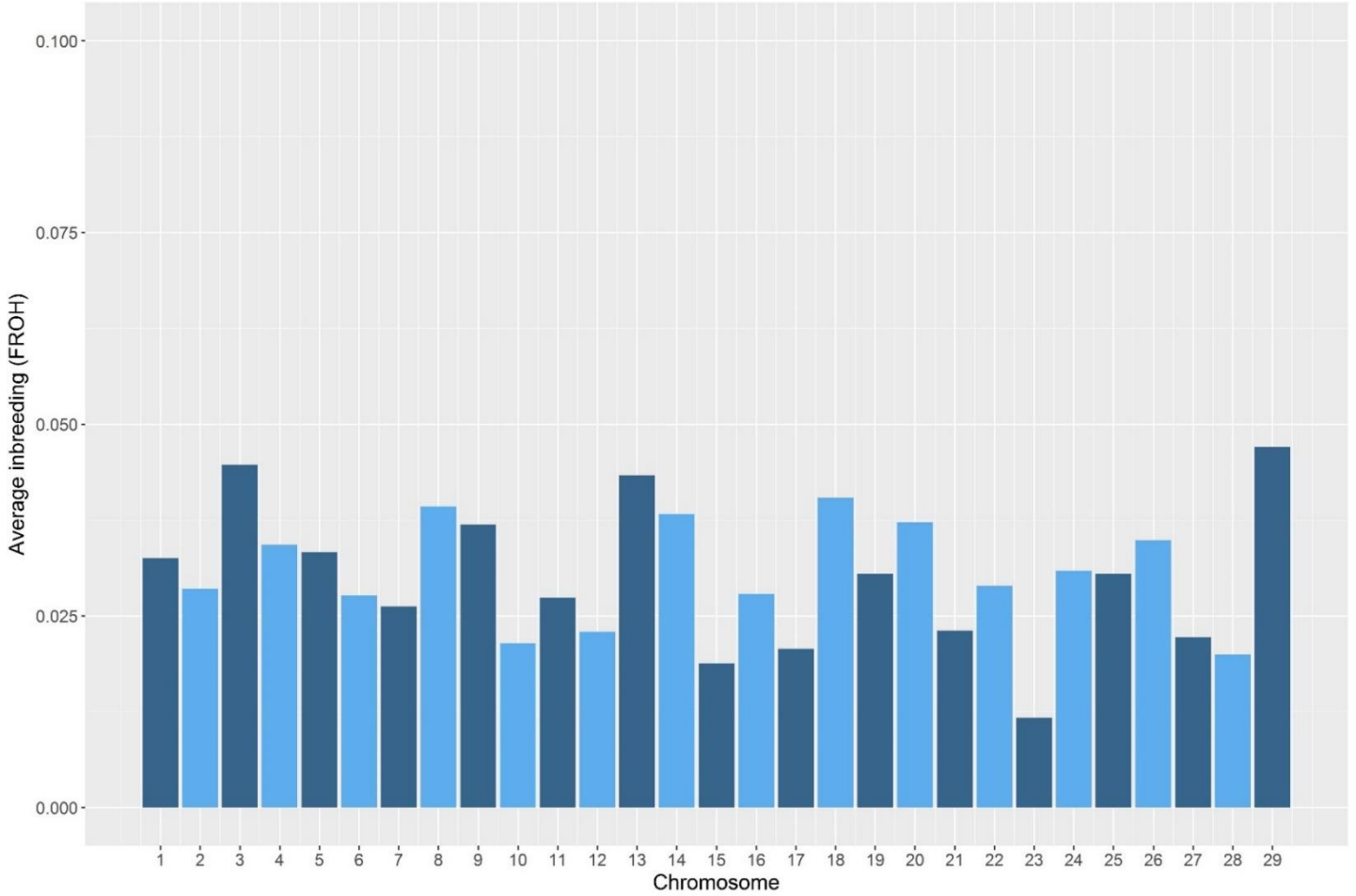
# Additional Files

**Additional file 1 Table 2.S1** Summary of the Angler cattle SNP data.

| BTA | No. of SNPs | BTA Length Covered by SNP (MB) | SNP Interval (Mb) | | | $r^2$ of adjacent SNP pairs (SD) |
|---|---|---|---|---|---|---|
| | | | Longest | Average | SD | |
| 1 | 2,875 | 158.229 | 0.567 | 0.055 | 0.047 | 0.176(0.231) |
| 2 | 2,333 | 138.223 | 0.895 | 0.059 | 0.061 | 0.173(0.234) |
| 3 | 2,174 | 121.403 | 0.668 | 0.056 | 0.050 | 0.173(0.228) |
| 4 | 2,152 | 120.642 | 0.449 | 0.056 | 0.045 | 0.158(0.216) |
| 5 | 1,847 | 121.176 | 0.752 | 0.066 | 0.062 | 0.160(0.221) |
| 6 | 2,201 | 129.996 | 7.121 | 0.059 | 0.160 | 0.166(0.225) |
| 7 | 1,933 | 112.629 | 1.227 | 0.058 | 0.061 | 0.165(0.217) |
| 8 | 2,086 | 113.367 | 0.429 | 0.054 | 0.042 | 0.154(0.204) |
| 9 | 1,743 | 105.588 | 0.667 | 0.061 | 0.056 | 0.170(0.229) |
| 10 | 1,843 | 104.215 | 1.962 | 0.057 | 0.074 | 0.149(0.205) |
| 11 | 1,914 | 107.043 | 0.823 | 0.056 | 0.050 | 0.161(0.220) |
| 12 | 1,469 | 91.092 | 3.082 | 0.062 | 0.107 | 0.164(0.227) |
| 13 | 1,515 | 84.149 | 0.615 | 0.055 | 0.048 | 0.194(0.241) |
| 14 | 1,532 | 85.006 | 0.556 | 0.055 | 0.047 | 0.175(0.225) |
| 15 | 1,480 | 85.257 | 0.663 | 0.057 | 0.050 | 0.146(0.208) |
| 16 | 1,425 | 81.341 | 1.361 | 0.057 | 0.063 | 0.157(0.217) |
| 17 | 1,375 | 74.998 | 0.835 | 0.054 | 0.056 | 0.144(0.201) |
| 18 | 1,156 | 65.979 | 0.967 | 0.057 | 0.054 | 0.162(0.216) |
| 19 | 1,182 | 64.007 | 0.675 | 0.054 | 0.048 | 0.147(0.205) |
| 20 | 1,377 | 71.794 | 0.559 | 0.052 | 0.042 | 0.169(0.226) |
| 21 | 1,178 | 71.137 | 1.216 | 0.060 | 0.069 | 0.183(0.239) |
| 22 | 1,095 | 61.378 | 0.466 | 0.056 | 0.046 | 0.168(0.238) |
| 23 | 939 | 52.285 | 0.523 | 0.056 | 0.047 | 0.139(0.196) |
| 24 | 1,091 | 62.686 | 0.431 | 0.057 | 0.046 | 0.165(0.230) |
| 25 | 859 | 42.851 | 0.283 | 0.050 | 0.037 | 0.157(0.213) |
| 26 | 942 | 51.680 | 0.401 | 0.054 | 0.043 | 0.158(0.224) |
| 27 | 854 | 45.633 | 0.819 | 0.053 | 0.054 | 0.132(0.196) |
| 28 | 833 | 46.235 | 0.345 | 0.055 | 0.044 | 0.138(0.198) |
| 29 | 928 | 51.503 | 1.180 | 0.055 | 0.055 | 0.150(0.206) |
| Total | 44,331 | 2,521.523 | 7.121 | 0.057 | 0.065 | 0.162(0.221) |

**Additional file 2 Table 2.S2.** The number of runs of homozygosity (ROH) in different categories detected in each chromosome.

| BTA | Total | ROH length category (Mb) | | | |
|---|---|---|---|---|---|
| | | 1-4 | 4-8 | 8-16 | >16 |
| 1 | 118 | 44 | 40 | 20 | 14 |
| 2 | 84 | 24 | 32 | 16 | 12 |
| 3 | 115 | 47 | 34 | 19 | 15 |
| 4 | 88 | 25 | 32 | 19 | 12 |
| 5 | 77 | 9 | 40 | 17 | 11 |
| 6 | 82 | 33 | 23 | 19 | 7 |
| 7 | 68 | 13 | 28 | 22 | 5 |
| 8 | 75 | 27 | 17 | 20 | 11 |
| 9 | 74 | 18 | 27 | 20 | 9 |
| 10 | 55 | 22 | 18 | 12 | 3 |
| 11 | 63 | 22 | 16 | 17 | 8 |
| 12 | 59 | 29 | 21 | 4 | 5 |
| 13 | 92 | 45 | 21 | 18 | 8 |
| 14 | 85 | 31 | 32 | 17 | 5 |
| 15 | 38 | 7 | 19 | 9 | 3 |
| 16 | 65 | 26 | 29 | 7 | 3 |
| 17 | 37 | 13 | 12 | 8 | 4 |
| 18 | 64 | 22 | 17 | 21 | 4 |
| 19 | 57 | 20 | 24 | 10 | 3 |
| 20 | 60 | 14 | 23 | 17 | 6 |
| 21 | 51 | 25 | 15 | 9 | 2 |
| 22 | 55 | 23 | 22 | 9 | 1 |
| 23 | 18 | 7 | 9 | 0 | 2 |
| 24 | 44 | 17 | 15 | 6 | 6 |
| 25 | 40 | 15 | 19 | 4 | 2 |
| 26 | 53 | 20 | 26 | 4 | 3 |
| 27 | 29 | 13 | 12 | 2 | 2 |
| 28 | 25 | 9 | 7 | 7 | 2 |
| 29 | 83 | 41 | 27 | 13 | 2 |
| Total | 1,854 | 661 | 657 | 366 | 170 |




**Additional file 3 Figure 2.S1.** Average inbreeding coefficient for each chromosome ($F_{ROH>1}$) for the Angler cattle. The average inbreeding coefficient for each chromosome was calculated as the proportion of autosome in runs of homozygosity (ROH) over the length of the BTA covered by the involved SNPs.

**Additional file 4 Table 2.S3.** Number of animals have SNP in the ROH in the corresponding chromosome

| | Number of animals have SNP in the ROH in the corresponding chromosome | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **BTA** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** |
| 1 | 0 | 0 | 115 | 309 | 355 | 497 | 520 | 428 | 274 | 217 | 95 | 10 | 39 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 116 | 212 | 86 | 416 | 404 | 388 | 267 | 160 | 104 | 153 | 26 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 11 | 117 | 170 | 203 | 229 | 223 | 203 | 236 | 241 | 109 | 154 | 93 | 59 | 63 | 38 | 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 124 | 126 | 256 | 310 | 232 | 182 | 310 | 120 | 40 | 94 | 149 | 78 | 34 | 11 | 33 | 52 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 117 | 79 | 196 | 142 | 177 | 309 | 228 | 207 | 172 | 102 | 41 | 66 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 47 | 296 | 391 | 284 | 324 | 99 | 202 | 147 | 139 | 110 | 110 | 52 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 28 | 162 | 192 | 339 | 269 | 257 | 207 | 129 | 51 | 81 | 58 | 58 | 72 | 14 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 51 | 97 | 305 | 238 | 111 | 151 | 190 | 438 | 279 | 207 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 0 | 18 | 0 | 75 | 247 | 320 | 346 | 162 | 136 | 66 | 123 | 72 | 102 | 40 | 4 | 31 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 1 | 275 | 232 | 208 | 459 | 212 | 254 | 71 | 67 | 18 | 46 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | 18 | 140 | 317 | 177 | 179 | 213 | 280 | 229 | 192 | 144 | 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | 18 | 170 | 105 | 153 | 311 | 328 | 132 | 80 | 95 | 19 | 1 | 33 | 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 0 | 0 | 0 | 25 | 72 | 147 | 170 | 215 | 247 | 272 | 183 | 69 | 62 | 30 | 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | 0 | 7 | 11 | 159 | 90 | 228 | 334 | 112 | 167 | 96 | 58 | 65 | 16 | 31 | 3 | 1 | 28 | 64 | 8 | 7 | 44 | 3 | 0 | 0 | 0 | 0 |
| 15 | 56 | 79 | 334 | 311 | 396 | 98 | 39 | 31 | 37 | 69 | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | 0 | 16 | 85 | 191 | 311 | 303 | 222 | 129 | 91 | 35 | 23 | 4 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17 | 0 | 126 | 194 | 447 | 228 | 103 | 97 | 67 | 53 | 7 | 46 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | 0 | 71 | 82 | 67 | 15 | 134 | 186 | 147 | 99 | 73 | 67 | 31 | 4 | 59 | 9 | 25 | 7 | 0 | 24 | 2 | 16 | 33 | 5 | 0 | 0 | 0 |
| 19 | 63 | 89 | 72 | 104 | 53 | 123 | 163 | 263 | 96 | 90 | 19 | 34 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | 0 | 4 | 41 | 225 | 304 | 220 | 72 | 84 | 67 | 21 | 55 | 76 | 85 | 17 | 19 | 7 | 44 | 4 | 21 | 3 | 8 | 0 | 0 | 0 | 0 | 0 |
| 21 | 10 | 55 | 101 | 354 | 187 | 200 | 86 | 99 | 42 | 39 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 22 | 0 | 0 | 47 | 136 | 247 | 191 | 171 | 111 | 59 | 85 | 48 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | 86 | 266 | 263 | 183 | 80 | 61 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 24 | 19 | 41 | 81 | 97 | 113 | 114 | 155 | 171 | 155 | 37 | 71 | 14 | 14 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25 | 0 | 31 | 104 | 52 | 93 | 157 | 143 | 84 | 86 | 54 | 7 | 6 | 35 | 6 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 26 | 0 | 12 | 64 | 188 | 161 | 136 | 125 | 84 | 73 | 8 | 3 | 2 | 0 | 10 | 3 | 2 | 1 | 6 | 0 | 7 | 10 | 11 | 0 | 3 | 22 | 11 |
| 27 | 0 | 142 | 18 | 179 | 132 | 178 | 77 | 48 | 26 | 1 | 53 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 28 | 32 | 98 | 85 | 139 | 175 | 182 | 81 | 34 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 29 | 0 | 0 | 65 | 21 | 53 | 56 | 54 | 81 | 75 | 125 | 114 | 83 | 27 | 13 | 69 | 12 | 27 | 27 | 26 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

# CHAPTER 3

# Novel optimum contribution selection methods accounting for conflicting objectives in breeding programs for livestock breeds with historical migration

Yu Wang[1*], Jörn Bennewitz[1], Robin Wellmann[1]

[1]Institute of Animal Science, University of Hohenheim, 70593 Stuttgart, Germany

[*]Corresponding author

Genetics Selection Evolution (2017) 49:45

doi: 10.1186/s12711-017-0320-7

# Abstract


## Background

Optimum contribution selection (OCS) is effective for increasing genetic gain, controlling the rate of inbreeding and enables maintenance of genetic diversity. However, this diversity may be caused by high migrant contributions (MC) in the population due to introgression of genetic material from other breeds, which can threaten the conservation of small local populations. Therefore, breeding objectives should not only focus on increasing genetic gains but also on maintaining genetic originality and diversity of native alleles. This study aimed at investigating whether OCS was improved by including MC and modified kinships that account for breed origin of alleles. Three objective functions were considered for minimizing kinship, minimizing MC and maximizing genetic gain in the offspring generation, and we investigated their effects on German Angler and Vorderwald cattle.

## Results

In most scenarios, the results were similar for Angler and Vorderwald cattle. A significant positive correlation between MC and estimated breeding values of the selection candidates was observed for both breeds, thus traditional OCS would increase MC. Optimization was performed under the condition that the rate of inbreeding did not exceed 1% and at least 30% of the maximum progress was achieved for all other criteria. Although traditional OCS provided the highest breeding values under restriction of classical kinship, the magnitude of MC in the progeny generation was not controlled. When MC were constrained or minimized, the kinship at native alleles increased compared to the reference scenario. Thus, in addition to constraining MC, constraining kinship at native alleles is required to ensure that native genetic diversity is maintained. When kinship at native alleles was constrained, the classical kinship was automatically lowered in most cases and more sires were selected. However, the average breeding value in the next generation was also lower than that obtained with traditional OCS.

## Conclusions

For local breeds with historical introgressions, current breeding programs should focus on increasing genetic gain and controlling inbreeding, as well as maintaining the genetic originality of the breeds and the diversity of native alleles via the inclusion of MC and kinship at native alleles in the OCS process.

# 3.1 Background

In recent decades, the widespread use of artificial insemination and other reproductive technologies has resulted in substantial genetic gains in livestock populations. However, another consequence is that only a limited number of animals with high estimated breeding values (EBV) have been intensively used in breeding programs, which can result in increasing rates of inbreeding to undesired levels. A high rate of inbreeding not only leads to considerable reduction in genetic variation but also more deleterious recessive alleles become homozygous, which may threaten the entire future of the population (Falconer and Mackay 1996). Thus, there is a conflict between maximizing genetic gain and managing the rate of inbreeding.

Crossbreeding has been demonstrated to be an efficient method to reduce the threat of inbreeding depression and increase the level of genetic diversity (Frankham *et al.* 2002). In addition, local breeds are often crossed with breeds of high economic value to improve performance. However, such introgressions of genetic material can be a threat for maintaining local breeds. Amador *et al.* (2011) confirmed that, after several generations without management, even a small introduction of foreign genetic material will rapidly disperse throughout the original population, and that this material is difficult to remove. Therefore, foreign introgressions present a large risk for the conservation of local breeds, which leads to a conflict in current breeding programs between increasing the contribution of foreign genetic material and conserving local breeds.

Optimum contribution selection (OCS) is a selection method that is effective at achieving a balance between rate of inbreeding and genetic gain. This selection process maximizes genetic gain in the next generation while constraining the rate of inbreeding via restriction of relatedness among offspring (Meuwissen 1997; Grundy *et al.* 1998; Woolliams *et al.* 2015). The superiority of OCS has been demonstrated with both simulated (Gourdine *et al.* 2012; Sánchez-Molano *et al.* 2016) and real data (Schierenbeck *et al.* 2011; Howard *et al.* 2014; Dagnachew and Meuwissen 2016). The objective function for OCS has been optimized using Lagrange multipliers (Meuwissen 1997; Eynard *et al.* 2016; Sánchez-Molano *et al.* 2016), evolutionary algorithms (Sørensen *et al.* 2006, 2008; Gourdine *et al.* 2012), and semidefinite programming algorithms (Pong-Wong and Woolliams 2007; Schierenbeck *et al.* 2011; Gómez-Romano *et al.* 2016). A similar related optimization problem was expressed as a mixed-integer quadratically constrained optimization problem and solved with branch-and-bound algorithms

(Mullin and Belotti 2016). In this paper, we applied the algorithm described in Pfaff (2014) for solving cone-constrained convex problems by using R package *optiSel*.

OCS is efficient for controlling the level of kinship among progeny and the rate of inbreeding in future generations and can ultimately maintain genetic diversity (Stachowicz *et al.* 2004; Ducro and Windig 2014; Gómez-Romano *et al.* 2016). However, a high level of genetic diversity can be achieved by a large genetic contribution from migrant breeds, which is undesirable for the conservation of local breeds, because it reduces their genetic uniqueness, as well as the genetic diversity between breeds (Bennewitz *et al.* 2008). Thus, conflicting objectives are observed with regards to maintaining genetic diversity and conserving genetic uniqueness of local small breeds with historical migrations.

Instead of focusing on genetic gain and rate of inbreeding only, a reasonable breeding objective would be to also include recovery of genetic originality by reducing migrant contributions (MC). The diversity of native alleles may also be important for conservation. Thus, to conserve breeds with historical migrations, Wellmann *et al.* (2012) recommended that approaches should not only constrain MC, but also aim at increasing the probability that alleles originating from native founders are not identical by descent (IBD).

Our aim was to investigate whether including MC and modified kinship matrices that account for breed origin of alleles as additional constraints in OCS can improve breeding programs in local breeds. Both conservation progress and genetic gain were evaluated. The following scenarios based on different objective functions were considered: (1) maximizing the diversity of native alleles while restricting MC and/or the average breeding value of the progeny generation at desired levels; (2) minimizing MC while restricting the loss of diversity of native alleles and/or the average breeding value of the progeny generation at desired levels; and (3) maximizing the average breeding value of the progeny generation while restricting MC and/or the loss of diversity of native alleles at desired levels. The traditional pedigree-based kinship was constrained in all optimization scenarios.

# 3.2 Methods

## 3.2.1 Data

Data from two local German cattle breeds, Angler and Vorderwald, were analyzed. The Angler breed is mainly located in the northern part of Germany and represents a dual-purpose breed, although the primary emphasis is on milk production. With the introduction of other breeds to improve milk yield, the Angler breed has experienced a considerable amount of migrant breed introgressions (Bennewitz and Meuwissen 2005). The Angler dataset was provided by the VIT (Vereinigte Informationssysteme Tierhaltung w.V., Verden), Germany. The Vorderwald breed is a dual-purpose breed located in the black forest region of southwest Germany. Similarly, due to their frequent crossing with high-yield breeds, the genetic originality of Vorderwald cattle has decreased dramatically (Hartwig *et al.* 2014, 2015). The Vorderwald dataset was provided by the Institute for Animal Breeding, Bavarian State Research Center for Agriculture in Grub, Germany. Both datasets consist of pedigrees with information on sex, breed, birth year and estimated breeding values for milk production obtained from routine genetic evaluations. Animals with an unknown pedigree born before 1970 were classified as purebred. Animals from other breeds and animals with an unknown pedigree born after 1970 were considered as migrants, although some may have purebred ancestors. The Angler dataset included 109,109 animals born between 1906 and 2015, of which 86,269 (79.1%) were classified as Angler. The Vorderwald dataset included 200,468 animals born between 1906 and 2010, of which 180,646 (90.1%) were classified as Vorderwald. MC for each animal was calculated and expressed as the proportion of migrant breed alleles based on pedigree information.

## 3.2.2 Selection candidates

Selection candidates were chosen among animals that were classified as purebred in the herdbook in order to compute their optimum contributions with different approaches. Sires that had progeny born in 2005 and 2006 were set as male selection candidates and selected males were mated to 1000 randomly chosen dams, which are called female selection candidates. For the Angler breed, 1199 selection candidates were available and 15,370 animals were involved in the pedigree that included all selection candidates and their ancestors. For the Vorderwald breed, 1123 selection candidates were available and 12,934 animals were involved in the

pedigree. For a better comparison of results between the two breeds, EBV were normalized across all selection candidates of each breed, with a mean of 0 and a standard deviation of 1.

### 3.2.3 Optimum contribution selection strategies

The output of the optimum contribution selection procedure is a vector **c** with individual genetic contributions. The genetic contribution $c_i$ of animal $i$ is the fraction of genes in the next generation that originate from this individual. Genetic contributions cannot be negative, i.e. $c_i \geq 0$, which is denoted as constraint (a) in the following. The total genetic contribution of each sex must be equal to 0.5 for diploid species, i.e. $\mathbf{c}'\mathbf{s} = 0.5$ and $\mathbf{c}'\mathbf{d} = 0.5$ (constraint b), where **s** and **d** are vectors of the indicators (0/1) of a candidate's sex. Because cows can produce only a limited number of calves, all female selection candidates were used for breeding and the genetic contributions were forced to be equal, i.e. $c_{d_1} = c_{d_2} = \cdots = c_{d_n}$ (constraint c). Thus, optimization was only performed for bulls. For male selection candidates, the number of offspring is not limited, thus the maximum genetic contribution is 0.5, *i.e.* $c_{s_i} \leq 0.5$. To calculate the proportion of sires with non-zero genetic contributions, a sire $i$ is considered to have a non-zero genetic contribution only if $c_{s_i} \geq 0.00025$ to account for possible numerical inaccuracies of the algorithm.

Four kinships that are involved in the calculation of the OCS procedure were applied. The diversity parameters described in Wellmann *et al.* (2012) are complementary to the kinships used here, i.e. these kinship values are equal to 1 minus the corresponding diversity denoted as $\varphi_A, \ldots, \varphi_D$ in Wellmann *et al.* (2012). The relevant derivations of the formulas for calculating the diversity parameters are provided in detail in Wellmann *et al.* (2012).

The classic kinship $f_A$ between individuals $i$ and $j$ (element of matrix $\mathbf{f_A}$), which describes the probability that two alleles, $X_i$ and $X_j$, at a locus that are randomly selected from individuals $i$ and $j$ are IBD (i.e. $\mathbf{f_A}(i,j) = \mathbf{P}\left(X_i \underbrace{=}_{\text{IBD}} X_j\right)$), was restricted in all scenarios. For breeds with historical migrations and foreign introgressions, Wellmann *et al.* (2012) proposed that the breed origin of the alleles should be considered to preserve the local breed. Thus, we considered different approaches that account for the origin of alleles, denoted as $f_B$, $f_C$ and $f_D$. Kinship matrix $\mathbf{f_B}$ contains the probabilities that two alleles randomly chosen from two individuals at a locus are IBD or that at least one allele is from a migrant breed ($\mathcal{M}$):

$$\mathbf{f_B}(i,j) = \mathbf{P}\left( X_i \underbrace{=}_{\text{IBD}} X_j \text{ or } X_i \in \boldsymbol{\mathcal{M}} \text{ or } X_j \in \boldsymbol{\mathcal{M}} \right).$$

Note that this is equal to the probability that both alleles are IBD and native plus the probability that at least one allele is from a migrant.

Kinship matrix $\mathbf{f_C}$ contains the probabilities that two alleles randomly chosen from two individuals at a locus are IBD or both alleles are from migrant breeds:

$$\mathbf{f_C}(i,j) = \mathbf{P}\left( X_i \underbrace{=}_{\text{IBD}} X_j \text{ or } X_i, X_j \in \boldsymbol{\mathcal{M}} \right).$$

This is equal to $\mathbf{f_B}(i,j) = \mathbf{f_C}(i,j) + \boldsymbol{P}(either\ X_i \in \boldsymbol{\mathcal{M}} \text{ or } X_j \in \boldsymbol{\mathcal{M}})$. The probability that at least one of the two randomly chosen alleles is from a migrant breed is higher than the probability that both are from migrant breeds. Thus, $\mathbf{f_B}$ is greater than $\mathbf{f_C}$. In general, $\mathbf{f_A} \leq \mathbf{f_C} \leq \mathbf{f_B}$ (element-wise). The kinship at native alleles $f_D$ is defined as the conditional probability that two alleles *X* and *Y* at a locus that are randomly chosen from the offspring population are IBD, given that both descended from native founders ($\boldsymbol{\mathcal{F}}$):

$$f_D(\mathbf{c}) = \mathbf{P}\left( X \underbrace{=}_{\text{IBD}} Y \mid X, Y \in \boldsymbol{\mathcal{F}} \right).$$

Note that this value says nothing about the kinship at loci that originate from migrants or about the MC. The mean kinships for the offspring generation are $\mathbf{c'f_A c}$, $\mathbf{c'f_B c}$ and $\mathbf{c'f_C c}$, respectively. Mean kinship $f_D$ in the offspring population was calculated as $f_D(\mathbf{c}) = 1 - \frac{1-\mathbf{c'f_B c}}{\mathbf{c'f_N c}}$, where $\mathbf{f_N}$ is a matrix containing the probabilities that both randomly chosen alleles at a locus originated from native founders.

Our aim was to identify the best method of accounting for the conflicting objectives of a breeding program, which are to increase breeding values, to maintain genetic diversity, and to maintain genetic originality of the breed. Since $1 - f_D(\mathbf{c}) = \mathbf{P}\left( X \underbrace{\neq}_{\mathbf{IBD}} Y \mid X, Y \in \boldsymbol{\mathcal{F}} \right)$ is the genetic diversity at native alleles, the constraint on $f_D$ is used to maintain or increase genetic diversity at native alleles and is a parameter of interest. Kinship $f_B$ and $f_C$ were considered because minimizing or constraining $f_D$ is in general not a convex problem, so minimizing $f_B$ and $f_C$ could result in lower $f_D$ values than minimizing $f_D$ itself.

In the different scenarios, an upper bound for MC ($\mathrm{ub.MC}$) and/or a lower bound for the average EBV ($\mathrm{lb.EBV}$) were set as additional constraints. The expectation of the average EBV in the next generation is $\mathbf{c}'\mathbf{EBV}$, where $\mathbf{EBV}$ is a vector of the EBV of each selection candidate. The expectation of the average MC of the next generation is $\mathbf{c}'\mathbf{MC}$, where $\mathbf{MC}$ is a vector of the MC of each selection candidate.

For all optimization problems, constraints a, b, and c were applied to limit the solution for $c_i$ to within a reasonable range. Solver "*cccp*" (Pfaff 2014), which was called from the R package *optiSel* (Wellmann 2017), was used to solve the optimization problems. This solver contains routines for solving cone constrained convex problems using interior-point methods that are partially ported from Python's CVXOPT and based on Nesterov-Todd scaling (Vandenberghe 2010). The solver uses a primal-dual path following algorithms for linear and quadratic cone constrained programming.

Scenarios were categorized based on three main objective functions: minimizing kinships, minimizing MC and maximizing genetic gain in the next generation. For minimizing kinships, three sub-scenarios were considered, which involved minimizing $f_B$, $f_C$ and $f_D$, respectively. Parameters $\mathrm{ub.}f_A$, $\mathrm{ub.}f_B$, $\mathrm{ub.}f_C$, $\mathrm{ub.}f_D$ and $\mathrm{ub.MC}$ were defined as the upper bound values of the corresponding parameters in the next generation, whereas $\mathrm{lb.EBV}$ was set as the lower bound of the mean EBV for the next generation. One or several of the following constraints were used to define the optimization problems for each breed:

$$\mathbf{c}'\mathbf{f_A}\mathbf{c} \leq \mathrm{ub.}f_A,$$

$$\mathbf{c}'\mathbf{f_B}\mathbf{c} \leq \mathrm{ub.}f_B,$$

$$\mathbf{c}'\mathbf{f_C}\mathbf{c} \leq \mathrm{ub.}f_C,$$

$$f_D(\mathbf{c}) \leq \mathrm{ub.}f_D,$$

$$\mathbf{c}'\mathbf{MC} \leq \mathrm{ub.MC},$$

$$\mathbf{c}'\mathbf{EBV} \geq \mathrm{lb.EBV}.$$

The OCS scenarios considered are listed in Table 3.1. The name of each optimization scenario consists of a prefix that indicates the objective function and a suffix that indicates the constraint settings. For example, scenario *maxEBV.A.B.MC* indicates a scenario that maximizes the average EBV in the next generation, while constraining $f_A$, $f_B$, and MC. The vector of genetic contributions for this scenario is denoted as $\mathbf{c}_{\text{maxEBV.A.B.MC}}$.

**Table 3.1** Names of the OCS scenarios based on different objective functions

| Objective function | Name of the scenario[1] |
|---|---|
| Minimizing $f_B$ | *minfB.A; minfB.A.MC; minfB.A.MC.EBV* |
| Minimizing $f_C$ | *minfC.A; minfC.A.MC; minfC.A.MC.EBV* |
| Minimizing $f_D$ | *minfD.A; minfD.A.MC; minfD.A.MC.EBV* |
| Minimizing MC | *minMC.A; minMC.A.EBV; minMC.A.B.EBV; minMC.A.C.EBV; minMC.A.D.EBV* |
| Maximizing EBV | *maxEBV.A; maxEBV.A.MC; maxEBV.A.B.MC; maxEBV.A.C.MC; maxEBV.A.D.MC* |

[1]The name of each optimization scenario consists of a prefix that indicates the objective function and a suffix that indicates the constraint settings. For example, scenario *minfB.A* indicates that the objective function is to minimize the average $f_B$ value in the following generation with a constraint on $f_A$.

Criteria for comparing scenarios included not only the result of the objective function, but also the other parameters obtained in the scenario, in particular EBV, MC, classic kinship, and kinship at native alleles. To evaluate the effectiveness of the OCS scenarios, the results were compared with the output from a reference scenario (*REF*) and the output from a truncation selection scenario (*TS*). In scenario *REF* all selection candidates were used as parents and had equal contributions to the offspring generation. For endangered breeds, an effective population size ($N_e$) of 50 is often considered as sufficient (Meuwissen 2009). Based on the equation in (Falconer and Mackay 1996), $\frac{1}{N_e} = \frac{1}{4*N_{sire}} + \frac{1}{4*N_{dam}}$, the 13 sires with the highest EBV were selected as male selection candidates in the *TS* scenario, and mated to the 1000 dams. All parents had equal contributions to the offspring generation in this scenario.

To ensure that optimal solutions exist in all scenarios for each breed, feasible threshold values must be set for the constraints. To restrict the rate of inbreeding, the upper bound ($\text{ub.}f_A$) was defined as follows. When $N_e$ is equal to 50, the rate of inbreeding ΔF, which can be calculated from $\Delta F = \frac{1}{2N_e}$, is 1% per generation. Based on this, the threshold for $f_A$ was calculated as $\text{ub.}f_A = \bar{f_A} + \left(1 - \bar{f_A}\right)\Delta F$, where $\bar{f_A}$ is the average kinship of the selection candidates.

To calculate the constraint setting for the other parameters, we used the results from the scenario that optimizes the corresponding parameter with restriction only on $f_A$ and the *REF* scenario, using the following calculations:

$$\text{ub.}\, f_B = \lambda \mathbf{c}'_{\text{minfB.A}} \mathbf{f_B} \mathbf{c}_{\text{minfB.A}} + (1-\lambda) \mathbf{c}'_{\text{REF}} \mathbf{f_B} \mathbf{c}_{\text{REF}},$$

$$\text{ub.}\, f_C = \lambda \mathbf{c}'_{\text{minfC.A}} \mathbf{f_C} \mathbf{c}_{\text{minfC.A}} + (1-\lambda) \mathbf{c}'_{\text{REF}} \mathbf{f_C} \mathbf{c}_{\text{REF}},$$

$$\text{ub.}\, f_D = \lambda \mathbf{f_D}(\mathbf{c}_{\text{minfD.A}}) + (1-\lambda) \mathbf{f_D}(\mathbf{c}_{\text{REF}}),$$

$$\text{ub.}\, \text{MC} = \lambda \mathbf{c}'_{\text{minMC.A}} \mathbf{MC} + (1-\lambda) \mathbf{c}'_{\text{REF}} \mathbf{MC},$$

$$\text{lb.}\, \text{EBV} = \lambda \mathbf{c}'_{\text{maxEBV.A}} \mathbf{EBV} + (1-\lambda) \mathbf{c}'_{\text{REF}} \mathbf{EBV},$$

where λ is a parameter that indicates the proportion of progress to be accomplished for each constrained parameter relative to the scenario with a restriction only on $f_A$. The value of λ can be determined by the breeding organization. A higher λ value indicates a stricter setting for all constraints. We set λ at 0.3 to ensure that optimized solutions were found for all scenarios and for both breeds. The specific values used for all constraints for each breed are in Additional file 1: Table 3.S1.

# 3.3 Results

Results of the basic statistical analyses for average kinship, MC and EBV of the parent generation are in Table 3.2 for both breeds. Average kinship $f_A$ was lower for the Angler population than for the Vorderwald population (0.020 vs. 0.025) but $f_B$ (0.910 vs. 0.853) and $f_C$ levels (0.488 vs. 0.381) were higher. On average, 69.5 and 60.7% of the genetic material of the Angler and Vorderwald cattle, respectively, originated from migrant breeds. Native effective population sizes of 86 and 49 were estimated from six previous generations for Angler and Vorderwald cattle, respectively. Native effective population size is a parameter that quantifies the decrease in native allele diversity and is defined in Wellmann *et al.* (2012). If the native effective size is high, then native allele diversity decreases slowly. Thus, the diversity of native alleles decreased more rapidly in Vorderwald cattle than in Angler cattle, whereas MC were higher in Angler cattle. Average EBV for both breeds were below the current population mean, which is 100 for Angler and 0 for Vorderwald because selection candidates

were sampled from old age cohorts. A positive correlation between EBV and MC was found for both breeds (Figure 3.1 and Figure 3.2).

**Table 3.2** Descriptive statistics for the active breeding population in the Angler and Vorderwald breeds.

| | Angler (N = 1199) | | Vorderwald (N = 1123) | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| $f_A$ | 0.020 | 0.027 | 0.025 | 0.027 |
| $f_B$ | 0.910 | 0.055 | 0.853 | 0.084 |
| $f_C$ | 0.488 | 0.123 | 0.381 | 0.128 |
| MC | 0.695 | 0.126 | 0.607 | 0.153 |
| EBV | 86.868 | 13.901 | -512.020 | 502.465 |

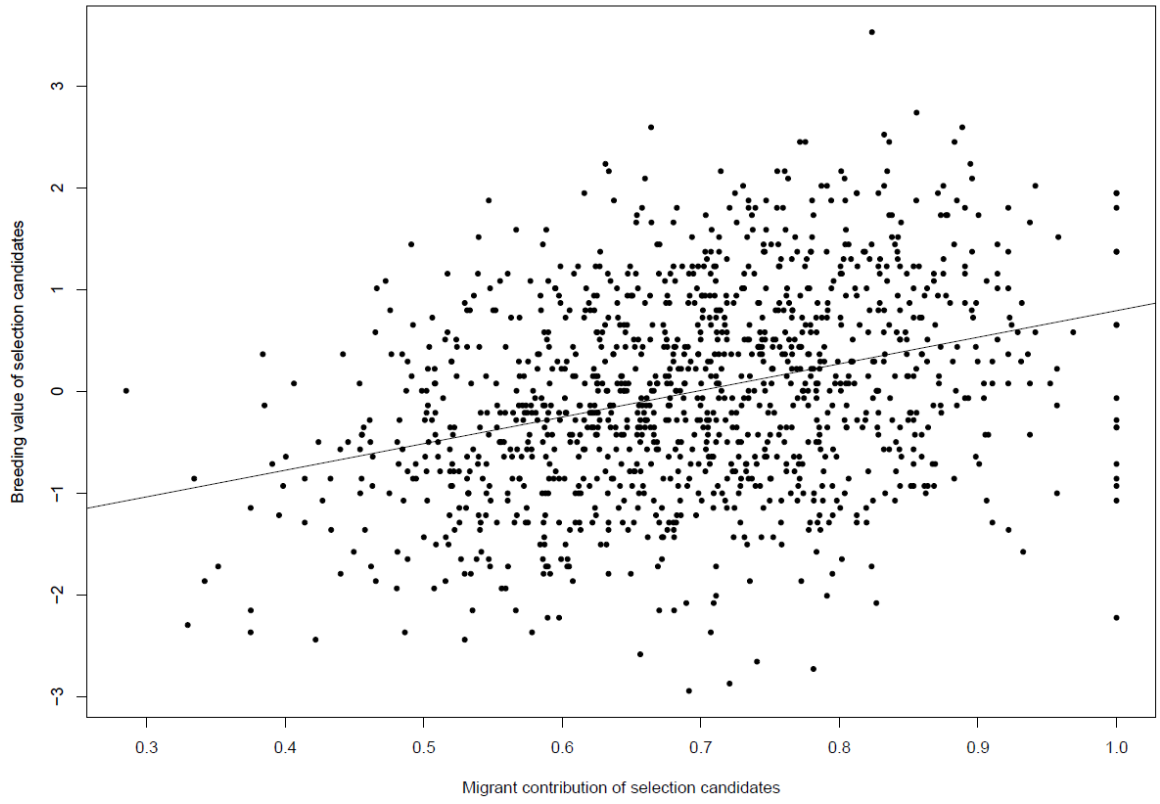



**Figure 3.1** Relationship between migrant contribution and the estimated breeding value of selection candidates in the Angler cattle population.

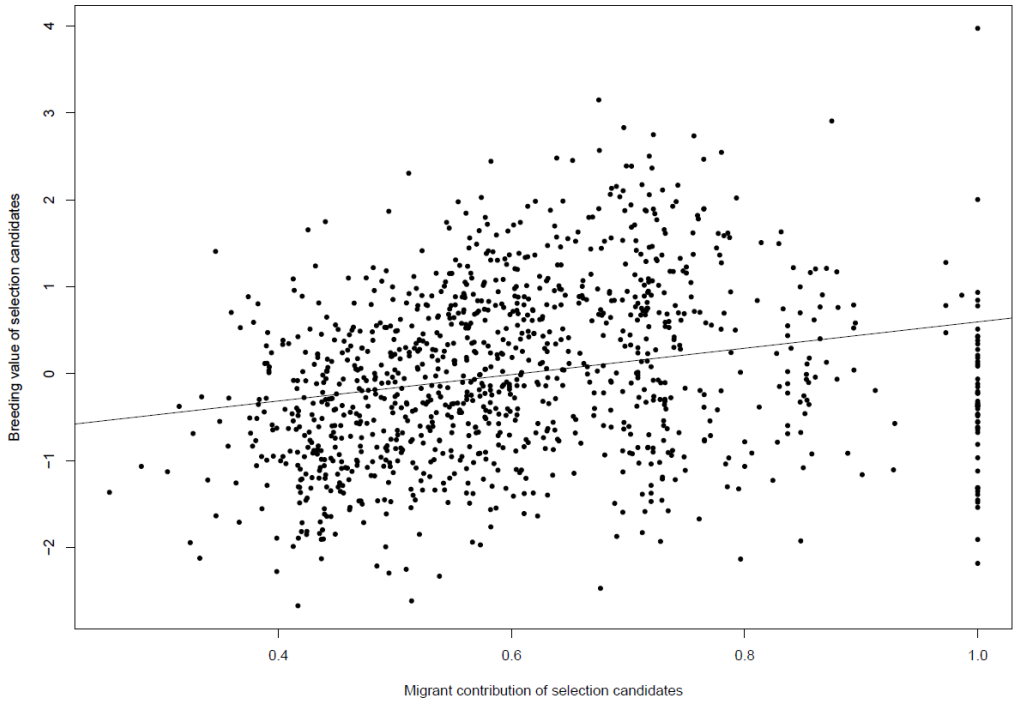



**Figure 3.2** Relationship between migrant contribution and the estimated breeding value of selection candidates in the Vordervald cattle population.

### 3.3.1 Minimizing average kinship

Genetic contributions of the selection candidates were optimized to minimize $f_B$, $f_C$ and $f_D$ with restrictions on MC and/or average EBV in the offspring generation for each breed, (see Tables 3.3, 3.4 and 3.5, respectively). Compared to the *REF* scenario, all OCS scenarios showed superior results for the optimized criteria as expected.

Table 3.3 shows the results obtained when minimizing $f_B$ in the offspring generation under the different constraints for each breed. The lowest $f_B$ for Angler cattle was 0.827 when the upper bound for $f_A$ in the next generation was set to 0.030. MC was lower than the constraint value setting (0.570 vs. 0.677). Thus, the minimum $f_B$ did not change after adding the constraint on MC (*minfB.A.MC*). When the restriction on average EBV was set to 0.516, the average kinship $f_B$ increased to 0.866, which was still lower than the $f_B$ obtained in the *REF* scenario (0.926). Similar results were obtained for Vorderwald cattle. When the upper bound for $f_A$ in the progeny generation was set to 0.035, the minimum $f_B$ level in the progeny generation was 0.789. Again, $f_B$ did not change after adding an upper bound for MC (0.528 vs. 0.582). $f_B$ increased to 0.813 when the EBV constraint was set to 0.550, although it was lower than the $f_B$ obtained in the *REF* scenario (0.852).

Results when minimizing $f_C$ were similar to minimizing $f_B$ (see Table 3.4). The $f_C$ of the progeny generation decreased to 0.345 for Angler cattle when the upper bound for $f_A$ was set to 0.030. When $f_C$ was minimized, MC decreased to a value lower than the constraint level setting (0.570 vs. 0.677). Thus, minimizing $f_C$ gave the same results for scenarios *minfC.A* and *minfC.A.MC*. After adding an EBV constraint of 0.516, $f_C$ increased to 0.404 but was lower than the $f_C$ obtained in the *REF* scenario (0.527). For Vorderwald cattle, the minimum average $f_C$ in the progeny generation was 0.300 when $f_A$ was restricted to 0.035, even after adding a higher constraint on MC (0.582 vs. 0.528). In scenario *minfC.A.MC.EBV*, $f_C$ reached 0.327 after adding an EBV constraint of 0.550, although this was lower than the $f_C$ obtained in the *REF* scenario (0.380).

**Table 3.3** Optimization of the genetic contributions when minimizing kinship $f_B$ with a restriction on migrant contribution and/or mean estimated breeding values

| **Scenario**[1] | **Parameter**[2] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **$f_A$** | **$f_B$**[3] | **$f_C$** | **$f_D$** | **MC** | **EBV** | **Selected**[4] | **SD of $c_s$**[5] |
| **Angler** | | | | | | | | |
| *REF* | 0.022 | 0.926 | 0.527 | 0.049 | 0.722 | 0.211 | - | - |
| *TS* | 0.031 | 0.939 | 0.565 | 0.067 | 0.722 | 1.184 | 0.065 | 0 |
| *minfB.A* | **0.030** | 0.827 | 0.345 | 0.082 | 0.570 | -0.295 | 0.106 | 0.012 |
| *minfB.A.MC* | **0.030** | 0.827 | 0.345 | 0.082 | ***0.570*** | -0.295 | 0.111 | 0.012 |
| *minfB.A.MC.EBV* | **0.030** | 0.866 | 0.404 | 0.083 | ***0.623*** | **0.516** | 0.081 | 0.012 |
| **Vorderwald** | | | | | | | | |
| *REF* | 0.030 | 0.852 | 0.380 | 0.072 | 0.605 | 0.287 | - | - |
| *TS* | 0.043 | 0.882 | 0.432 | 0.093 | 0.645 | 1.161 | 0.106 | 0 |
| *minfB.A* | **0.035** | 0.789 | 0.300 | 0.074 | 0.528 | -0.111 | 0.260 | 0.011 |
| *minfB.A.MC* | **0.035** | 0.789 | 0.300 | 0.074 | ***0.528*** | -0.111 | 0.260 | 0.011 |
| *minfB.A.MC.EBV* | **0.035** | 0.813 | 0.327 | 0.075 | ***0.555*** | **0.550** | 0.228 | 0.010 |

[1]The name of each optimization scenario consists of a prefix that indicates the objective function and a suffix that indicates the constraint settings

[2]The parameter used as a constraint is marked in bold in the scenario. Italic values indicate that the actual value obtained does not reach the limit of the corresponding constraint (value higher than the lower limit or lower than the upper limit)

[3]Objective function

[4]Proportion of selected sires with non-zero genetic contributions; a $c_{s_i}$ value lower than 0.00025 is treated as zero

[5]Standard deviation of the genetic contributions of all male selection candidates

**Table 3.4** Optimization of the genetic contribution when minimizing kinship $f_C$ with a restriction on migrant contribution and/or mean estimated breeding values

| **Scenario**[1] | **Parameter**[2] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **$f_A$** | **$f_B$** | **$f_C$**[3] | **$f_D$** | **MC** | **EBV** | **Selected**[4] | **SD of $c_s$**[5] |
| **Angler** | | | | | | | | |
| *REF* | 0.022 | 0.926 | 0.527 | 0.049 | 0.722 | 0.211 | - | - |
| *TS* | 0.031 | 0.939 | 0.565 | 0.067 | 0.722 | 1.184 | 0.065 | 0 |
| *minfC.A* | **0.030** | 0.827 | 0.345 | 0.082 | 0.570 | -0.299 | 0.111 | 0.012 |
| *minfC.A.MC* | **0.030** | 0.827 | 0.345 | 0.082 | ***0.570*** | -0.299 | 0.111 | 0.012 |
| *minfC.A.MC.EBV* | **0.030** | 0.866 | 0.404 | 0.083 | ***0.623*** | **0.516** | 0.091 | 0.012 |
| **Vorderwald** | | | | | | | | |
| *REF* | 0.030 | 0.852 | 0.380 | 0.072 | 0.605 | 0.287 | - | - |
| *TS* | 0.043 | 0.882 | 0.432 | 0.093 | 0.645 | 1.161 | 0.106 | 0 |
| *minfC.A* | **0.035** | 0.789 | 0.300 | 0.074 | 0.528 | -0.109 | 0.276 | 0.010 |
| *minfC.A.MC* | **0.035** | 0.789 | 0.300 | 0.074 | ***0.528*** | -0.109 | 0.276 | 0.010 |
| *minfC.A.MC.EBV* | **0.035** | 0.813 | 0.327 | 0.075 | ***0.555*** | **0.550** | 0.228 | 0.010 |

[1]The name of each optimization scenario consists of a prefix indicating the objective function and a suffix indicating the constraint settings

[2]The parameter used as a constraint is marked in bold in the scenario. Italic values show that the actual value obtained does not reach the limit of the corresponding constraint in this scenario (value higher than the lower limit or lower than the upper limit)

[3]Objective function

[4]Proportion of selected sires with non-zero genetic contributions; a $c_{s_i}$ value lower than 0.00025 is treated as zero

[5]Standard deviation of the genetic contributions of all male selection candidates

When the kinship at native alleles, $f_D$ , was minimized, the average kinship $f_A$ was automatically lowered in most cases (Table 3.5); in Angler cattle, $f_A$ reached 0.020, which was lower than the constraint level (0.030). In this case, the minimum $f_D$ was 0.040. When MC was restricted to 0.677, the minimum $f_D$ increased to 0.044. When an EBV constraint of 0.516 was added, the minimum $f_D$ increased to 0.047, which was still lower than the $f_D$ obtained in the *REF* scenario (0.049). For Vorderwald cattle, when $f_A$ was restricted to 0.035 in the progeny generation, the lowest $f_D$ was 0.057. When the maximum MC was set to 0.582, $f_D$ increased to 0.058. When adding an EBV constraint of 0.550, the lowest $f_D$ was 0.064, which was still lower than the $f_D$ obtained in the *REF* scenario (0.072).

### 3.3.2 Minimizing migrant contribution

Table 3.6 shows the results of minimizing MC under various constraints. When $f_A$ was restricted to 0.030 in the progeny generation for Angler cattle, MC was equal to 0.570. When constraining the EBV to at least 0.516, MC in scenario *minMC.A.EBV* increased to 0.622 and $f_B$ and $f_C$ were lower than their constraint settings (0.866 vs. 0.896 and 0.404 vs. 0.472, respectively). Thus, adding constraints for $f_B$ or $f_C$ did not change the results. When the upper bound for $f_D$ was set to 0.046, MC increased to 0.683, which was less than that achieved in the *REF* scenario (0.722). Results were similar for Vorderwald cattle. The minimum MC achieved in the next generation was 0.527 when the upper bound for $f_A$ was 0.035. When the lower bound for EBV was set to 0.550, the minimal MC increased to 0.555. Adding a lower constraint for $f_B$ (0.813 vs. 0.833) or $f_C$ (0.327 vs. 0.356) did not change results. When the upper bound for $f_D$ was set to 0.067 as an additional constraint, the minimum MC was 0.571, which was less than that obtained in the *REF* scenario (0.605).

**Table 3.5** Optimization of genetic contribution when minimizing kinship $f_D$ with restriction on migrant contribution and/or mean estimated breeding values

| **Scenario**[1] | **Parameter**[2] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **$f_A$** | **$f_B$** | **$f_C$** | **$f_D$**[3] | **MC** | **EBV** | **Selected**[4] | **SD of $c_s$**[5] |
| **Angler** | | | | | | | | |
| *REF* | 0.022 | 0.926 | 0.527 | 0.049 | 0.722 | 0.211 | - | - |
| *TS* | 0.031 | 0.939 | 0.565 | 0.067 | 0.722 | 1.184 | 0.065 | 0 |
| *minfD.A* | ***0.020*** | 0.954 | 0.614 | 0.040 | 0.782 | 0.078 | 0.434 | 0.009 |
| *minfD.A.MC* | ***0.019*** | 0.899 | 0.464 | 0.044 | **0.677** | 0.090 | 0.414 | 0.004 |
| *minfD.A.MC.EBV* | ***0.020*** | 0.899 | 0.464 | 0.047 | **0.677** | **0.516** | 0.333 | 0.005 |
| **Vorderwald** | | | | | | | | |
| *REF* | 0.030 | 0.852 | 0.380 | 0.072 | 0.605 | 0.287 | - | - |
| *TS* | 0.043 | 0.882 | 0.432 | 0.093 | 0.645 | 1.161 | 0.106 | 0 |
| *minfD.A* | **0.035** | 0.895 | 0.456 | 0.057 | 0.669 | 0.759 | 0.398 | 0.015 |
| *minfD.A.MC* | ***0.027*** | 0.833 | 0.352 | 0.058 | **0.582** | 0.145 | 0.472 | 0.006 |
| *minfD.A.MC.EBV* | ***0.029*** | 0.833 | 0.353 | 0.064 | **0.582** | **0.550** | 0.358 | 0.007 |

[1]The name of each optimization scenario consists of a prefix indicating the objective function and a suffix indicating the constraint settings

[2]The parameter used as a constraint is marked in bold in the scenario. Italic values show that the actual value obtained does not reach the limit of the corresponding constraint in this scenario (value higher than the lower limit or lower than the upper limit)

[3]Objective function

[4]Proportion of selected sires with non-zero genetic contributions; a $c_{s_i}$ value lower than 0.00025 is treated as zero

[5]Standard deviation of the genetic contributions of all male selection candidates

**Table 3.6** Optimization of the genetic contribution when minimizing the migrant contribution with restricted kinship and/or mean estimated breeding values

| Scenario[1] | Parameter[2] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $f_A$ | $f_B$ | $f_C$ | $f_D$ | MC[3] | EBV | Selected[4] | SD of $c_s$[5] |
| **Angler** | | | | | | | | |
| *REF* | 0.022 | 0.926 | 0.527 | 0.049 | 0.722 | 0.211 | - | - |
| *TS* | 0.031 | 0.939 | 0.565 | 0.067 | 0.722 | 1.184 | 0.065 | 0 |
| *minMC.A* | **0.030** | 0.827 | 0.345 | 0.083 | 0.570 | -0.289 | 0.106 | 0.012 |
| *minMC.A.EBV* | **0.030** | 0.866 | 0.404 | 0.084 | 0.622 | **0.516** | 0.091 | 0.012 |
| *minMC.A.B.EBV* | **0.030** | ***0.866*** | 0.404 | 0.084 | 0.622 | **0.516** | 0.091 | 0.012 |
| *minMC.A.C.EBV* | **0.030** | 0.866 | ***0.404*** | 0.084 | 0.622 | **0.516** | 0.091 | 0.012 |
| *minMC.A.D.EBV* | ***0.020*** | 0.903 | 0.472 | **0.046** | 0.683 | **0.516** | 0.342 | 0.005 |
| **Vorderwald** | | | | | | | | |
| *REF* | 0.030 | 0.852 | 0.380 | 0.072 | 0.605 | 0.287 | - | - |
| *TS* | 0.043 | 0.882 | 0.432 | 0.093 | 0.645 | 1.161 | 0.106 | 0 |
| *minMC.A* | **0.035** | 0.789 | 0.300 | 0.074 | 0.527 | -0.111 | 0.276 | 0.011 |
| *minMC.A.EBV* | **0.035** | 0.813 | 0.327 | 0.075 | 0.555 | **0.550** | 0.220 | 0.010 |
| *minMC.A.B.EBV* | **0.035** | ***0.813*** | 0.327 | 0.075 | 0.555 | **0.550** | 0.220 | 0.010 |
| *minMC.A.C.EBV* | **0.035** | 0.813 | ***0.327*** | 0.075 | 0.555 | **0.550** | 0.211 | 0.010 |
| *minMC.A.D.EBV* | ***0.031*** | 0.825 | 0.342 | **0.067** | 0.571 | **0.550** | 0.317 | 0.008 |

[1]The name of each optimization scenario consists of a prefix indicating the objective function and a suffix indicating the constraint settings

[2]The parameter used as a constraint is marked in bold in the scenario. Italic values show that the actual value obtained does not reach the limit of the corresponding constraint in this scenario (value higher than the lower limit or lower than the upper limit)

[3]Objective function

[4]Proportion of selected sires with non-zero genetic contributions; a $c_{s_i}$ value lower than 0.00025 is treated as zero

[5]Standard deviation of the genetic contributions of all male selection candidates

### 3.3.3 Maximizing the average EBV

Results for maximizing the average EBV in the progeny generation under various constraints are in Table 3.7. For both breeds, the *REF* scenario achieved the lowest average EBV in the offspring generation. This value was not zero because male and female selection candidates had different mean EBV. For Angler cattle, scenario *maxEBV.A* achieved a higher EBV (1.226 vs. 1.184) than the TS scenario, although the average kinship $f_A$ was restricted (0.030 vs. 0.031). The average EBV decreased when adding the MC restriction, and $f_B$ and $f_C$ decreased to a level lower than their upper bound settings. Restricting $f_D$ also lowered $f_A$. The EBV dropped to its lowest value of 0.449 when restricting $f_A$, $f_D$ and MC, although this was still around twice that obtained in the *REF* scenario (0.211). Similar results were observed for the Vorderwald cattle population. Scenario *maxEBV.A* achieved a similar EBV as the *TS* scenario (1.164 vs. 1.161) but the average kinship $f_A$ was much lower (0.035 vs. 0.043). When adding restrictions on $f_D$ and MC, the maximum EBV decreased to 0.636, which was more than twice that obtained in the REF scenario (0.287).

The number of selected sires with non-zero genetic contributions was calculated in each scenario, as well as the standard deviation of the genetic contribution of all male selection candidates. Among all scenarios, TS selected the smallest number of sires. Adding a constraint on $f_D$ resulted in all cases in more selected sires and a lower standard deviation.

**Table 3.7** Optimization of the genetic contribution when maximizing the breeding value with restricted kinship and/or mean estimated migrant contributions

| Scenario[1] | Parameter[2] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $f_A$ | $f_B$ | $f_C$ | $f_D$ | MC | EBV[3] | Selected[4] | SD of $c_s$[5] |
| **Angler** | | | | | | | | |
| *REF* | 0.022 | 0.926 | 0.527 | 0.049 | 0.722 | 0.211 | - | - |
| *TS* | 0.031 | 0.939 | 0.565 | 0.067 | 0.722 | 1.184 | 0.065 | 0 |
| *maxEBV.A* | **0.030** | 0.937 | 0.560 | 0.082 | 0.743 | 1.226 | 0.085 | 0.012 |
| *maxEBV.A.MC* | **0.030** | 0.901 | 0.471 | 0.082 | **0.677** | 0.979 | 0.070 | 0.012 |
| *maxEBV.A.B.MC* | **0.030** | ***0.893*** | 0.454 | 0.082 | ***0.664*** | 0.884 | 0.075 | 0.012 |
| *maxEBV.A.C.MC* | **0.030** | 0.901 | ***0.471*** | 0.082 | **0.677** | 0.979 | 0.070 | 0.012 |
| *maxEBV.A.D.MC* | ***0.020*** | 0.899 | 0.464 | **0.046** | **0.677** | 0.449 | 0.347 | 0.005 |
| **Vorderwald** | | | | | | | | |
| *REF* | 0.030 | 0.852 | 0.380 | 0.072 | 0.605 | 0.287 | - | - |
| *TS* | 0.043 | 0.882 | 0.432 | 0.093 | 0.645 | 1.161 | 0.106 | 0 |
| *maxEBV.A* | **0.035** | 0.895 | 0.456 | 0.077 | 0.666 | 1.164 | 0.203 | 0.013 |
| *maxEBV.A.MC* | **0.035** | 0.835 | 0.357 | 0.079 | **0.582** | 0.812 | 0.220 | 0.011 |
| *maxEBV.A.B.MC* | **0.035** | ***0.832*** | 0.353 | 0.078 | ***0.579*** | 0.787 | 0.220 | 0.011 |
| *maxEBV.A.C.MC* | **0.035** | 0.835 | **0.356** | 0.078 | ***0.581*** | 0.808 | 0.220 | 0.011 |
| *maxEBV.A.D.MC* | ***0.031*** | 0.834 | 0.354 | **0.067** | **0.582** | 0.636 | 0.317 | 0.008 |

[1]The name of each optimization scenario consists of a prefix indicating the objective function and a suffix indicating the constraint settings

[2]The parameter used as a constraint is marked in bold in the scenario. Italic values show that the actual value obtained does not reach the limit of the corresponding constraint in this scenario (value higher than the lower limit or lower than the upper limit)

[3]Objective function

[4]Proportion of selected sires with non-zero genetic contributions; a $c_{s_i}$ value lower than 0.00025 is treated as zero

[5]Standard deviation of the genetic contributions of all male selection candidates

# 3.4 Discussion

For the breeding schemes of the two breeds considered in this study, two conflicts must be addressed: (1) the conflict between increasing genetic gain while managing inbreeding and (2) the conflict between maintaining genetic diversity while controlling loss of genetic uniqueness. The purpose of this study was to determine whether OCS with additional constraints that involve modified kinship matrices and MC was more efficient at conserving genetic diversity and originality while also ensuring genetic improvement than traditional OCS. Using data on German Angler and Vorderwald cattle, various scenarios were compared. Both breeds have been frequently crossed with high-yielding breeds to improve performance. We found that diversity of native alleles decreased more rapidly in Vorderwald cattle than in Angler cattle, whereas MC was higher in Angler cattle. The consequences of the scenarios were similar for both breeds. Compared to traditional OCS, constraining kinship $f_D$ and MC promoted recovery of genetic originality in the breeds and diversity of native alleles but reduced response to selection.

Traditional OCS achieved the highest average EBV in the progeny generation among all scenarios with a restriction on rate of inbreeding, which, in our study, is represented by scenario *maxEBV.A*. Compared to the *TS* scenario, average EBV was higher in the traditional OCS scenario for both breeds, while the average relatedness was lower. Probably, the average EBV in *TS* was smaller because the *TS* scenario assumed equal contributions for selected sires, whereas OCS optimizes their contributions. Because MC and EBV were positively correlated, traditional OCS increased the average MC, which is undesirable when the aim is to conserve the genetic originality of local breeds.

## 3.4.1 Different kinship estimates

Both $f_B$ and $f_C$ take probabilities of IBD and probabilities of alleles originating from migrant breeds into account, i.e. they account for both level of inbreeding and level of genetic originality. Although theoretically, MC affects $f_B$ more than $f_C$, results from minimizing $f_B$ and $f_C$ were almost identical for the two breeds considered. Wellmann *et al.* (2012) reported a larger difference between these two methods, which is probably because contributions of both sexes were optimized in their work. Minimizing neither $f_B$ nor $f_C$ reduced the kinship at native alleles,

$f_D$, thus these two criteria are not an alternative for controlling the kinship at native alleles directly. Results from minimizing $f_B$ and $f_C$ were very similar to the results from minimizing MC. Hence, instead of minimizing or constraining $f_B$ or $f_C$, it is recommended to control MC. To control the diversity at native alleles, $f_D$ must be constrained or minimized directly, although this optimization problem may be not convex. However, because minimizing $f_D$ did not reduce MC, a constraint on MC is needed for all optimizations that involve $f_D$. Minimizing $f_D$ is different from minimizing $f_A$ with an additional constraint on MC because minimizing $f_A$ resulted in a larger $f_D$ than minimizing $f_D$ when MC is constrained to the same level (results not shown). Similarly, when including kinship $f_D$ as an additional constraint in the OCS, the level of kinship $f_A$ decreased in all scenarios. Thus, if $f_D$ is constrained, then MC must be constrained as well and the constraint for $f_A$ can be omitted.

Among all the scenarios, *TS* used the smallest number of sires and resulted in the highest average genetic contribution of selected sires. Including kinship $f_D$ as an additional constraint in the OCS scenarios resulted in a larger number of selected sires than including $f_B$ or $f_C$. Therefore, including $f_D$ is an efficient method to avoid overuse of sires with high EBV and limits the rate of inbreeding in the long run. Compared with the inclusion of $f_B$ or $f_C$, inclusion of $f_D$ resulted in a lower average EBV in the progeny generation, depending on the constraint level setting. In most cases, OC was negatively correlated with MC and positively correlated with the average EBV, as illustrated in Additional file 2 Table 3.S2, which represents a desirable result for future selection and breeding programs.

Scenarios with optimizations of both male and female contributions were also evaluated (results not shown), using the same calculation methods to obtain the constraint value settings. For all scenarios and both breeds, the constraint settings were stricter than in the scenarios that optimized male contributions. The performance of all scenarios improved when both male and female selection were optimized, which is consistent with Sánchez-Molano *et al.* (2016), who used OCS to improve fitness and productivity traits. To achieve these improvements, however, additional reproductive techniques must be applied due to the limited reproduction rate of female animals.

### 3.4.2 Considering the migrant contribution

Previous OCS approaches for maximizing genetic gain while limiting rate of inbreeding did not consider MC. Introgression of migrant breed alleles must be managed to maintain genetic uniqueness and conserve local breeds. As expected, the average EBV obtained with and without MC as a constraint showed that controlling MC restricts increases in genetic gain. Interestingly, kinship at native alleles increased compared to the *REF* scenario when MC was constrained or minimized. Hence, the individuals with the lowest MC may not carry some native alleles that are still present in individuals with higher MC. Thus, in this case, constraining $f_D$ is required to ensure that native genetic diversity is maintained.

However, maximum genetic gains can only be achieved by allowing for the introgression of foreign genetic material. Therefore, the two main purposes in a breeding program, i.e. conserving local breeds and improving genetic gain, are contradictory and must be balanced by the breeding organization. In this study, we set the proportion of breeding progress to be achieved at $\lambda = 0.3$ to determine the constraint level required for achieving optimal solutions for both breeds. Depending on the situation, the breeding organization could select an appropriate value of $\lambda$ to emphasize conservation of local breeds or genetic improvement, thus facilitating both purposes.

### 3.4.3 Future improvements

Because of advances in molecular genetics, genome-wide dense marker genotype data are increasingly available, even for some endangered breeds and have shown promise in capturing genetic variation due to Mendelian sampling (Avendaño *et al.* 2004). The application of genomic data provides a more accurate method of calculating relationships between individuals compared with the use of estimates from pedigree data (Sonesson *et al.* 2012). Breeding values estimated by genomic approaches are also more accurate and show more within-family variation compared with breeding values estimated via traditional approaches (Hill 2013). Furthermore, compared to the use of pedigree kinship, the use of genomic kinship is substantially more efficient in maintaining genetic diversity when optimizing genetic contributions (Clark *et al.* 2013; Gómez-Romano *et al.* 2016; Sánchez-Molano *et al.* 2016). Moreover, new methods to estimate kinship at native alleles, *i.e.* $f_D$, can be developed based on genomic data and the use of genomic data may enable estimation of MC for selection candidates without using pedigree data.

# 3.5 Conclusions

Maintaining genetic originality is essential for conserving local breeds. It was shown that using an OCS approach as developed in this study can effectively maintain the diversity of native alleles and genetic originality, while ensuring genetic improvement. The most promising approach involved the inclusion of additional constraints for migrant contributions and kinship at native alleles $f_D$. When a constraint for $f_D$ was included, the classical kinship $f_A$ in the offspring was lower than the constraint level, so the constraint on $f_A$ could be removed. More sires were selected when $f_D$ was constrained than when $f_D$ was not constrained and the standard deviation of the contributions was lower, i.e., the optimum contributions of the selected sires were more similar.

## Authors' contributions

YW did the statistical analyses and wrote the first draft of the paper. RW developed the OCS approaches and wrote the software. RW and JB initiated and oversaw the project and refined the manuscript. All authors read and approve the final manuscript.

## Acknowledgements

The study was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG). The authors thank the VIT (Vereinigte Informationssysteme Tierhaltung w.V., Verden), Germany for providing the Angler data and the Institute of Animal Breeding of the Bavarian State Research Center for Agriculture for providing the Vorderwald data. In addition, we wish to thank the editor JCM Dekkers as well as three anonymous reviewers for their constructive suggestions to improve the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

# Reference

Amador C., Toro M. Á., Fernández J., 2011 Removing exogenous information using pedigree data. Conserv. Genet. 12: 1565–1573.

Avendaño S., Woolliams J. A., Villanueva B., 2004 Mendelian sampling terms as a selective advantage in optimum breeding schemes with restrictions on the rate of inbreeding. Genet. Res. 83: 55–64.

Bennewitz J., Meuwissen T. H. E., 2005 Estimation of extinction probabilities of five German cattle breeds by population viability analysis. J. Dairy Sci. 88: 2949–2961.

Bennewitz J., Simianer H., Meuwissen T. H. E., 2008 Investigations on merging breeds in genetic conservation schemes. J. Dairy Sci. 91: 2512–2519.

Clark S. A., Kinghorn B. P., Hickey J. M., van der Werf J., 2013 The effect of genomic information on optimal contribution selection in livestock breeding programs. Genet. Sel. Evol. 45: 44.

Dagnachew B. S., Meuwissen T. H. E., 2016 A fast Newton–Raphson based iterative algorithm for large scale optimal contribution selection. Genet. Sel. Evol. 48:70.

Ducro B. J., Windig J. J., 2014 Genetic diversity and measures to reduce inbreeding in Friesian Horses. In: *Proceedings of the 10$^{th}$ world congress on genetics applied to livestock production*: 17–22 August 2014; Vancouver

Eynard S. E., Windig J. J., Hiemstra S. J., Calus M. P. L., 2016 Whole-genome sequence data uncover loss of genetic diversity due to selection. Genet. Sel. Evol. 48: 33.

Falconer D. S., Mackay T. F. C., 1996 *Introduction to quantitative genetics*. 4$^{th}$ ed Harlow: Pearson Education Limited.

Frankham R., Briscoe D. A., Ballou J. D., 2002 *Introduction to Conservation Genetics*. 2$^{nd}$ ed Cambridge: Cambridge University Press.

Gómez-Romano F., Villanueva B., Fernández J., Woolliams J. A., Pong-Wong R., 2016 The use of genomic coancestry matrices in the optimisation of contributions to maintain genetic diversity at specific regions of the genome. Genet. Sel. Evol. 48: 2.

Gourdine J. L., Sørensen A. C., Rydhmer L., 2012 There is room for selection in a small local pig breed when using optimum contribution selection: A simulation study. J. Anim. Sci. 90: 76–84.

Grundy B., Villanueva B., Woolliams J. A., 1998 Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. Genet. Res. (Camb). 72: 159–168.

Hartwig S., Wellmann R., Hamann H., Bennewitz J., 2014 The contribution of migrant breeds to the genetic gain of beef traits of German Vorderwald and Hinterwald cattle. J. Anim. Breed. Genet. 131: 496–503.

Hartwig S., Wellmann R., Emmerling R., Hamann H., Bennewitz J., 2015 Short communication: Importance of introgression for milk traits in the German Vorderwald and Hinterwald cattle. J. Dairy Sci. 98: 2033–2038.

Howard D. M., Pong-Wong R., Knap P. W., Kremer V. D., Woolliams J. A., 2014 The structural impact of implementing optimal contribution selection in a commercial pig breeding population. In: *Proceedings of the 10$^{th}$ world congress on genetics applied to livestock production*: 17-22 August 2014; Vancouver

Meuwissen T. H. E., 1997 Maximizing the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75: 2575–2583.

Meuwissen T. H. E., 2009 Genetic management of small populations: A review. Acta Agr Scand A-AN 59: 71–79.

Mullin T. J., Belotti P., 2016 Using branch-and-bound algorithms to optimize selection of a fixed-size breeding population under a relatedness constraint. Tree Genet. Genomes 12:4.

Pfaff B., 2014 The R package *cccp*: Design for solving cone constrained convex programs. R Financ.: http://www.pfaffikus.de/files/conf/rif/rif2014.pdf.

Pong-Wong R., Woolliams J. A., 2007 Optimisation of contribution of candidate parents to maximise genetic gain and restricting inbreeding using semidefinite programming. Genet. Sel. Evol. 39: 3–25.

Sánchez-Molano E., Pong-Wong R., Banos G., 2016 Genomic-based optimum contribution in conservation and genetic improvement programs with antagonistic fitness and productivity traits. Front. Genet. 7: 25.

Schierenbeck S., Pimentel E. C. G., Tietze M., Körte J., Reents R., *et al.*, 2011 Controlling inbreeding and maximizing genetic gain using semi-definite programming with pedigree-based and genomic relationships. J. Dairy Sci. 94: 6143–52.

Skotte L., Korneliussen T. S., Albrechtsen A., 2013 Estimating individual admixture proportions from next generation sequencing data. Genetics 195: 693–702.

Sonesson A. K., Woolliams J. A., Meuwissen T. H. E., 2012 Genomic selection requires genomic control of inbreeding. Genet. Sel. Evol. 44: 27.

Sørensen M. K., Sørensen A. C., Borchersen S., Berg P., 2006 Consequences of using EVA software as a tool for optimal genetic contribution selection in Danish Holstein. In: *Proceedings of the 8th world congress on genetics applied to livestock production*: 13-18 August 2006; Belo Horizonte

Sørensen M. K., Sørensen A. C., Baumung R., Borchersen S., Berg P., 2008 Optimal genetic contribution selection in Danish Holstein depends on pedigree quality. Livest. Sci. 118: 212–222.

Stachowicz K., Sørensen A. C., Berg P., 2004 Optimum contribution selection conserves genetic diversity better than random selection in small populations with overlapping generations. In: *Proceedings of the 55th annual meeting of the European Association for Animal Production (EAAP)*: 5–9 September 2004; Bled.

Vandenberghe L., 2010 The CVXOPT linear and quadratic cone program solvers. http://abel.ee.ucla.edu/cvxopt/documentation/coneprog.pdf.

Wellmann R., Hartwig S., Bennewitz J., 2012 Optimum contribution selection for conserved populations with historic migration. Genet. Sel. Evol. 44: 34.

Wellmann R., 2017 *optiSel*: optimum contribution selection and population genetics. R package version 0.7.1. 2017. https://cran.r-project.org/web/packages/optiSel/optiSel.pdf.

Woolliams J. A., Berg P., Dagnachew B. S., Meuwissen T. H. E., 2015 Genetic contributions and their optimization. J. Anim. Breed. Genet. 132: 89–99.

# Additional Files

**Additional File 1 Table 3.S1** Threshold settings for all parameters for the Angler and Vorderwald populations

| | **Angler** | | | **Vorderwald** | | |
|---|---|---|---|---|---|---|
| | **$Parameter_{REF}$[1]** | **Limit[2]** | **Constraint** | **$Parameter_{REF}$[1]** | **Limit[2]** | **Constraint** |
| $f_A$ | - | - | 0.030 | - | - | 0.035 |
| $f_B$ | 0.926 | 0.827 | 0.896[3] | 0.852 | 0.789 | 0.833 |
| $f_C$ | 0.527 | 0.345 | 0.472 | 0.380 | 0.300 | 0.356 |
| $f_D$ | 0.049 | 0.040 | 0.046 | 0.072 | 0.057 | 0.067 |
| MC | 0.722 | 0.570 | 0.677 | 0.605 | 0.527 | 0.582 |
| EBV | 0.211 | 1.226 | 0.516 | 0.287 | 1.164 | 0.550 |

[1]Output obtained from the reference scenario (*REF*)

[2]Upper or lower limit of the parameter in the scenario of the corresponding objective function with the constraint $f_A$

[3] Constraint value for $f_B$ is calculated as $ub.f_B = \lambda \mathbf{c}'_{minfB.A}\mathbf{f_B}\mathbf{c}_{minfB.A} + (1-\lambda)\mathbf{c}'_{REF}\mathbf{f_B}\mathbf{c}_{REF}$ , In this study, we set $\lambda$ to 0.3. Therefore, for Angler cattle, $ub.f_B$=0.3*0.827+(1-0.3)*0.926=0.896. The other following constraints settings were similarly calculated except for constraint $f_A$.

**Additional File 2 Table 3.S2** Correlation between OC and EBV and between OC and MC.

| | **Angler** | | **Vorderwald** | |
|---|---|---|---|---|
| **Scenario** | **Cor(OC,EBV)[1]** | **Cor(OC,MC)[2]** | **Cor(OC,EBV)[1]** | **Cor(OC,MC)[2]** |
| *minfB.A* | -0.217 | -0.573 | -0.323 | -0.521 |
| *minfB.A.MC* | -0.217 | -0.573 | -0.323 | -0.521 |
| *minfB.A.EBV.MC* | 0.130 | -0.375 | 0.216 | -0.341 |
| *minfC.A* | -0.219 | -0.574 | -0.325 | -0.526 |
| *minfC.A.MC* | -0.219 | -0.574 | -0.325 | -0.526 |
| *minfC.A.EBV.MC* | 0.131 | -0.378 | 0.217 | -0.344 |
| *minfD.A* | -0.077 | 0.301 | 0.276 | 0.310 |
| *minfD.A.MC* | -0.153 | -0.508 | -0.188 | -0.255 |
| *minfD.A.EBV.MC* | 0.324 | -0.428 | 0.308 | -0.225 |
| *minMC.A* | -0.213 | -0.569 | -0.322 | -0.518 |
| *minMC.A.EBV* | 0.128 | -0.371 | 0.217 | -0.343 |
| *minMC.A.B.EBV* | 0.128 | -0.371 | 0.217 | -0.343 |
| *minMC.A.C.EBV* | 0.128 | -0.371 | 0.217 | -0.343 |
| *minMC.A.D.EBV* | 0.341 | -0.394 | 0.272 | -0.290 |
| *maxEBV.A* | 0.427 | 0.076 | 0.563 | 0.326 |
| *maxEBV.A.MC* | 0.322 | -0.169 | 0.420 | -0.154 |
| *maxEBV.A.B.MC* | 0.285 | -0.219 | 0.401 | -0.175 |
| *maxEBV.A.C.MC* | 0.322 | -0.169 | 0.419 | -0.158 |
| *maxEBV.A.D.MC* | 0.268 | -0.454 | 0.365 | -0.201 |

[1]Correlation between the genetic contributions and estimated breeding values of all male selection candidates in the corresponding scenario

[2]Correlation between the genetic contributions and migrant contributions of all male selection candidates in the corresponding scenario

# CHAPTER 4

# Long-term impact of optimum contribution selection strategies on local livestock breeds with historical introgression using the example of German Angler cattle

Yu Wang[1*], Dierck Segelke[2], Reiner Emmerling[3], Jörn Bennewitz[1], Robin Wellmann[1]

[1]Institute of Animal Science, University of Hohenheim, 70593 Stuttgart, Germany

[2]Vereinigte Informationssysteme Tierhaltung w.V., Heideweg 1, 27283 Verden, Germany

[3]Institute of Animal Breeding, Bavarian State Research Centre for Agriculture, 85586 Grub, Germany

[*]Corresponding author

G3: Genes, Genomes, Genetics: vol. 7 no. 12 4009-4018

doi; 10.1534/g3.117.300272

# Abstract

The long-term performance of different selection strategies was evaluated via simulation using the example of a local cattle breed, the German Angler cattle. Different optimum contribution selection (OCS) approaches to maximize genetic gain were compared to a reference scenario without selection and truncation selection. The kinships and migrant contribution (MC) were estimated from genomic data. Truncation selection achieved the highest genetic gain but decreased diversity considerably at native alleles. It also caused the highest increase in MCs. Traditional OCS, which only constrains kinship, achieved almost the same genetic gain but also caused a small increase of MC and remarkably reduced the diversity at native alleles. When MC was required not to increase and the increase of kinship at native alleles was restricted, the migrant contribution levels and the diversity at native alleles were well managed, and the genetic gain was only slightly reduced. However, genetic progress was substantially lower in the scenario that aimed to recover the original genetic background. Truncation selection and traditional OCS both reduce the genetic originality of breeds with historical introgression. The inclusion of MC and kinship at native alleles as additional constraints in OCS showed great potential for conservation. Recovery of the original genetic background is possible but requires many generations of selection and reduces the genetic progress in performance traits. Hence, constraining MCs at their current values can be recommended to avoid further reduction of genetic originality.

**Keywords:** optimum contribution selection, conservation, genetic gain, migrant contribution, runs of homozygosity

# 4.1 Background

Crossbreeding can have positive and negative consequences for managed livestock populations. The introgressive hybridization of breeds with high economic value is common to improve performance. Furthermore, the gene flow between populations can counteract the loss of genetic diversity and avoid inbreeding depression. However, it is possible that persistent introgression of genetic material causes breeds to become genetically extinct. For the management of local breeds with historical introgression, three conflicts have to be addressed, *i.e.* the conflict between increasing genetic gain while managing the inbreeding level, the conflict between maintaining genetic diversity while controlling the loss of genetic uniqueness, and the conflict between increasing genetic gain while recovering the original genetic background. The traditional approach of optimum contribution selection (traditional OCS) provides a solution to solve the first problem. It aims to maximize genetic gain while controlling the rate of inbreeding by optimizing the genetic contribution of each selection candidate to the next generation (Meuwissen 1997; Grundy *et al.* 1998; Woolliams *et al.* 2015). However, traditional OCS cannot solve the other problems. For breeds with historical introgression, although OCS is efficient for controlling the level of kinship and maintaining genetic diversity (Eynard *et al.* 2016), this diversity may be caused by a large proportion of genetic contributions from other breeds. Additionally, although OCS is efficient in increasing genetic gain, this genetic gain may be achieved by sustained introgression with high-yielding breeds. The reduction of genetic uniqueness due to high migrant contributions (MCs) and the reduction of diversity at native alleles is a risk for the conservation of the genetic background of the breed. Apart from focusing on high breeding values, reducing MCs to recover genetic originality could also be included as a breeding objective. Advanced OCS approaches could effectively maintain native allele diversity and genetic originality, while ensuring genetic improvement by including MC and kinship at native alleles (Wellmann *et al.* 2012) as additional constraints, which has been shown for OCS based on pedigree information (Wang *et al.* 2017).

High-density marker panels of single-nucleotide polymorphisms (SNPs) allow us to obtain more accurate estimates of kinship than pedigrees, as it is common for pedigrees to contain errors (Ron *et al.* 1996). In addition, genotype-based kinship reflects the actual relatedness between two individuals, whereas pedigree-based estimates are only expectations (Visscher *et al.* 2006). Furthermore, genotype-based kinship is able to capture the relationships due to

distant common ancestors that pedigree-based estimates fail to reflect. Thus, using genotype-based kinship is more efficient than using a pedigree-based approach for the conservation of local breeds (Toro *et al.* 2014; Mészáros *et al.* 2015), especially for the removal of the undesired genetic materials (Amador *et al.* 2013). Currently, there are three approaches to estimate kinships from genome-wide SNPs. The first one is the molecular kinship; *i.e.*, the proportion of SNPs that are identical by state (IBS) (Eding and Meuwissen 2001; Caballero and Toro 2002). The second one is the genomic covariance between individuals computed from gene contents (VanRaden 2008; Yang *et al.* 2010). The third one is the segment-based kinship computed from shared haplotype segments, which are also known as runs of homozygosity (ROH) (Gusev *et al.* 2009; de Cara *et al.* 2013; Rodríguez-Ramilo *et al.* 2015; Gómez-Romano *et al.* 2016). Both molecular kinship and genomic relationship matrices have the disadvantage of being biased due to the preselection of markers included in the SNP panel (Nielsen 2000; McTavish and Hillis 2015). Moreover, increasing genetic diversity by reducing average molecular kinship drives allele frequencies towards 0.5 and increases the frequency of rare deleterious alleles. Thus, it accumulates deleterious variants in the genome and may reduce the fitness of the population (de Cara *et al.* 2011, 2013). The use of segment-based kinship has been shown to provide a good compromise between maintaining diversity and fitness levels in populations. The estimate based on segments reflects recent identity by descent (IBD) rather than identity by state (Keller *et al.* 2011). In this study, the segment-based kinship will be used in the optimization process.

Genomic information can also be used to estimate the breed composition of an individual (Frkonja *et al.* 2012). Software packages for predicting breed composition are usually based on either hidden Markov model clustering algorithms or maximum likelihood procedures (Pritchard *et al.* 2000; Alexander *et al.* 2009; Baran *et al.* 2012). Such analysis can be carried out by Admixture (Alexander *et al.* 2009) or Structure (Pritchard *et al.* 2000), where individuals are assumed to be unrelated, and linkage disequilibrium (LD) is not taken into account. Another approach is to assign haplotype segments to the breeds in which they have maximum frequency, which is carried out by *optiSel* (Wellmann 2017).

The objective of this study was to evaluate the long-term performance of different genomic OCS strategies, using the example of a local cattle breed, by simulating several subsequent generations. The scenarios were compared not only with respect to the genetic gain but also with respect to parameters measuring genetic diversity and genetic uniqueness.

# 4.2 Material and methods

## 4.2.1 Data

The dataset consisted of genotype information of 889 individuals belonging to five cattle breeds: 268 Angler, 200 Fleckvieh, 200 Holstein-Friesian, 200 Red Holstein and 21 Norwegian Red. The targeted breed in this study is the Angler cattle, which is a dual-purpose cattle breed with an emphasis on milk production. It is mainly located in the northern part of Germany (Bennewitz and Meuwissen 2005). The reference breeds, which include animals from the non-targeted breeds, were only used for the identification of native haplotype segments in Angler cattle. Two hundred Fleckvieh animals were genotyped with the Illumina BovineHD BeadChip (HD), and the remaining animals were genotyped with the Illumina BovineSNP50 BeadChip (50K) with standard quality control parameters. SNPs that were not available for all breeds were discarded. Finally, 23,448 autosomal SNPs were used for the analysis. Haplotypes were phased for all breeds jointly as part of a larger dataset, and missing genotypes were imputed using BEAGLE software (Browning and Browning 2007). To visualize the relationship between Angler and the other four breeds, principal component analysis (PCA) (Price *et al.* 2006) was performed on the SNP genotypes using PLINK 1.9 software (Chang *et al.* 2015).

## 4.2.2 Simulation

The simulations comprised two parts. First, a base population ($G_0$) was generated from the phased genotypes of the Angler cattle. Second, this base population was managed for the following 10 nonoverlapping generations in accordance with the respective scenario.

The base population $G_0$, consisting of 1000 simulated individuals, was generated from genotypes of 131 Angler bulls and 137 Angler cows based on a random sampling of gametes. The animals from other breeds which were used to identify native segments remained the same for each generation. The selection process started from generation $G_0$. For all scenarios, the optimum genetic contribution of each selection candidate ($c_i$) to the next generation was calculated for each generation. The corresponding number of offspring of each parent was generated, and mates were allocated randomly. Offspring received haplotypes from their parents via Mendelian inheritance, allowing recombination to occur according to the length of

the chromosomes; *i.e.*, one crossover occurs on average on a chromosome of size 1 Morgan (Weng *et al.* 2014). For all generations, the population size remained 1000 (500 males and 500 females). For each scenario, five replicates were simulated, and the results presented are averages over replicates.

A total of 1500 SNPs were sampled randomly without replacement to become quantitative trait loci (QTL). The QTL effects were sampled from a γ distribution with a shape parameter of 0.4 (Meuwissen *et al.* 2001) and standardized afterward. The effect of each QTL had a 50% chance of being positive or negative. The highest positive QTL effects were assigned to SNPs that were more frequent in the reference breeds than in Angler cattle. Hence the mean breeding value in Angler was lower than the mean breeding value in the reference breeds that were used for introgression.

The simulated true breeding value (TBV) of animal j was calculated as the sum of all QTL effects:

$$\mathrm{TBV_j} = \sum_{k=1}^{n_{QTL}} a_k \cdot Q_{kj},$$

where $n_{QTL} = 1500$ is the number of QTL, $a_k$ is the additive effect of QTL k, and $Q_{kj}$ is the QTL genotype of individual j at locus k. The genotypes were coded as 0, 1, or 2, as the number of copies of the alternative allele. For each individual, an estimated breeding value (EBV) for total merit with the reliability of 0.75 was simulated as:

$$\mathrm{EBV_j} = \mu_{EBV} + r^2(\mathrm{TBV_j} - \mu_{EBV}) + E_j$$

where $\mu_{EBV}$ is the mean of the breeding values of the corresponding generation, $E_j$ is a residual term sampled from a normal distribution with mean 0 and variance $\sigma_E^2 = r^2(1 - r^2)\sigma_{TBV}^2$.

### 4.2.3 Migrant contribution, kinships and diversity parameters

For calculating the kinship matrices and migrant contribution, the origin of each marker had to be determined for each haplotype from the breed of interest. A haplotype was classified to be native in Angler cattle at a particular marker position if the frequency of the segment containing the marker was sufficiently low in all reference breeds. Only haplotype segments consisting of ≧20 consecutive markers and a minimum length of 2.5 MB were considered. A marker was classified to be native in Angler if the frequency of the segment containing the marker was

<0.01 in all reference breeds. The migrant contribution of each individual was calculated as the proportion of its genome that was not classified to be native. The mathematical definitions can be found in the Appendix. For identification of the origin of markers, the R package *optiSel* (Wellmann 2017) was used.

Two SNP-based kinship parameters were considered, which are denoted as $f_{SEG}$ and $f_{SEG|N}$. Kinship $f_{SEG}$ between individual i and individual j (element of the matrix $\mathbf{f_{SEG}}$) is the probability that two alleles taken from a random position from randomly chosen haplotypes of both individuals belong to a shared segment, which is in accordance with de Cara *et al.* (2013). The mean kinship $\mathbf{f_{SEG}}$ for the offspring generation is estimated as $\mathbf{c'f_{SEG}c}$, where $\mathbf{c}$ is the vector of optimum genetic contributions of all selection candidates. In addition, average kinships among different breeds were calculated from a segment-based kinship matrix that included individuals from all breeds.

For breeds with historical introgression, Wellmann *et al.* (2012) proposed that kinship at native alleles should be restricted to preserve local breeds. The kinship $f_{SEG|N}$ is the conditional probability that two alleles taken at random from the population belong to a shared segment, given that they are native. For the computation of the segment based kinship $f_{SEG}$ and the kinship at native alleles $f_{SEG|N}$ we used R package *optiSel* (Wellmann 2017). The corresponding pedigree-based kinships were referred to as $f_A$ and $f_D$ in Wang *et al.* (2017). The mathematical definitions can be found in the Appendix.

Three additional genetic parameters were calculated to evaluate the level of genetic diversity of Angler cattle, *i.e.* the average observed heterozygosity $H_O$, the variance of the true breeding values ($\sigma^2_{TBV}$) and the genic variance ($\sigma^2_A$). The observed heterozygosity quantifies the amount of genetic variation due to polymorphic loci, which is an important parameter of estimating genetic variation within a population (Gregorius 1978). We calculated the $H_O$ of each generation in each scenario with software PLINK 1.9 (Chang *et al.* 2015). The genic variance was calculated as

$$\sigma^2_A = \sum_{m=1}^{n_{QTL}} 2p_m(1-p_m)a^2_m$$

where $n_{QTL} = 1{,}500$ is the number of QTL, $p_m$ is the allele frequency at locus m and $a_m$ is the additive effect of QTL m (Falconer and Mackay 1996).

### 4.2.4 Optimization scenarios

Except for the reference scenario, the objective of all the other scenarios was to maximize the genetic gain of the following generation, so the objective function was $\mathbf{c}'\mathbf{EBV}$, where $\mathbf{EBV}$ is a vector of the EBVs of all selection candidates. Three OCS scenarios were considered and compared to two non-OCS scenarios, *i.e.* a reference scenario without selection and a truncation selection scenario.

***Reference Scenario (REF)***

In this scenario, all animals were used as parents and each selection candidate had two offspring. Thus, no selection or optimization was done. The effective population size ($N_e$) is 2000 thus the increase of kinship is negligible.

***Truncation selection (TS)***

Maintenance of an effective population size of 100 was envisaged, as recommended in Meuwissen (2009). Calculated from $\frac{1}{N_e} \approx \frac{1}{4N_{Sire}} + \frac{1}{4N_{dam}}$ (Falconer and Mackay 1996), 26 bulls with the highest EBVs were selected for breeding to create the following generation by truncation selection. In this scenario, all selected bulls had equal contributions to the offspring. Note that the effective size in this scenario is expected to deviate slightly from 100 because the formula does not take into account how the individuals with highest breeding values are related.

***Traditional OCS method (OCS-I)***

To restrict the rate of inbreeding, the upper bound of kinship $f_{SEG}$ was defined as follows. Since the targeted effective population size was $N_e$=100, the desired rate of inbreeding, which can be calculated from $\Delta F = \frac{1}{2N_e}$ (Falconer and Mackay 1996), was 0.5% per generation. The threshold for $f_{SEG}$ of generation $t+1$ was calculated as:

$$\text{ub.}\, f_{SEG_{t+1}} = \overline{f_{SEG_t}} + \left(1 - \overline{f_{SEG_t}}\right) \Delta F,$$

where $\overline{f_{SEG_t}}$ is the average kinship of the population in generation $t$.

***OCS with constraint on kinship $f_{SEG}$, kinship $f_{SEG|N}$ and MC (OCS-II)***

The constraint of kinship $f_{SEG}$ was the same as in the OCS-I scenario. Additionally, constraints on conditional kinship $f_{SEG|N}$ and MC were applied. The upper bound threshold for $f_{SEG|N}$ in generation $t + 1$ was calculated as:

$$\text{ub.}\, f_{SEG|N_{t+1}} = \overline{f_{SEG|N_t}} + \left(1 - \overline{f_{SEG|N_t}}\right) \Delta F$$

where $\overline{f_{SEG|N_t}}$ is the mean kinship at native alleles of the population at generation $t$. Additionally, we required that for each generation, the average level for the estimated migrant contribution does not exceed the average level in the base generation $G_0$ ($\overline{MC_{G_0}}$).

***OCS with constraint on kinship $f_{SEG}$, kinship $f_{SEG|N}$ and reduced level of MC (OCS-III)***

The upper bounds of kinship $f_{SEG}$ and of kinship $f_{SEG|N}$ were the same as in Scenario *OCS-II* for each generation. Additionally, in this scenario, we required that the MC level estimated from haplotypes decreased 3% per generation.

Several reasonable conditions were made for all scenarios. The genetic contribution of a selection candidate ($c_i$), expressed as the proportion of genetic material originating from this individual in the next generation, was assumed to be non-negative ($c_i \geq 0$). In diploid species, each sex group contributes half of the genes to the gene pool. Thus, the sum of genetic contribution of all selection candidates of a sex was 0.5; *i.e.,* $\mathbf{c's} = 0.5$ and $\mathbf{c'd} = 0.5$, where **s** and **d** are vectors for indicators of a candidate's sex. For all OCS scenarios, optimization was done only for males. All females were assumed to have equal numbers of offspring. All 500 males were used as selection candidates, which reflects a breeding program with genomic selection in which a substantial number of the bull calves are genotyped.

The specific values for each constraint are shown in Table 4.S1 [See Additional file 1 Table 4.S1]. Solver "*cccp*" (Pfaff 2014), which was called from the R package *optiSel* version 0.9.1(Wellmann 2017), was used to solve the optimization problems. Five replicates per scenario were simulated and the results presented are averages across replicates.

### 4.2.5 Data availability

The data used in this study are available as supplementary files. File 4.S1 contains SNP ID numbers and locations. File 4.S2 contains simulated genotypes for each individual of Angler base generation $G_0$.

## 4.3 Results

### 4.3.1 Analysis of base generation ($G_0$)

The simulated base generation reflects the structure of the genotyped animals well. PCA plots of both populations were almost identical (Figure 4.1 and Figure 4.S1 [See Additional file 2 Figure 4.S1]). The first and second principal components (PC1 and PC2) separated animals in the simulated base generation according to their breed (Figure 4.1). PC1 explained 23.81% of the total variance and distinguished Fleckvieh from the other four breeds. Angler was separated from the Holstein family by PC2, which explained 15.85% of the total variance. Overlap existed between Holstein-Friesian and Red Holstein since Red Holstein is known to be a subpopulation of Holstein-Friesian.

Relationships within and between breeds are shown in Table 4.1. The smallest average kinship within a breed was found in Angler (0.048). This shows that Angler has a higher genetic diversity and lower inbreeding than the other breeds. Angler had a close relationship with Red Holstein (0.039) and Holstein-Friesian (0.037), a moderate close relationship with Norwegian Red (0.017) and a distant relationship with Fleckvieh (0.004). This is in agreement with the estimated genetic contributions the Angler has from other breeds, which were 0.448 from Holstein-Friesian and Red Holstein, 0.152 from Norwegian Red, and 0.021 from Fleckvieh (data not shown).

A basic statistical analysis of the simulated true breeding values of animals based on each breed group is presented in Table 4.2. Angler and Fleckvieh had relatively low average TBVs, with a mean of 0.560 and 0.367, respectively. Holstein-Friesian, Norwegian Red and Red Holstein had relatively high average TBVs, with a mean of 2.160, 2.390 and 2.431, respectively (as desired). As shown in Figure 4.S2 [See Additional file 3 Figure 4.S2], there was a positive correlation between MC and TBV in the base population of Angler cattle.

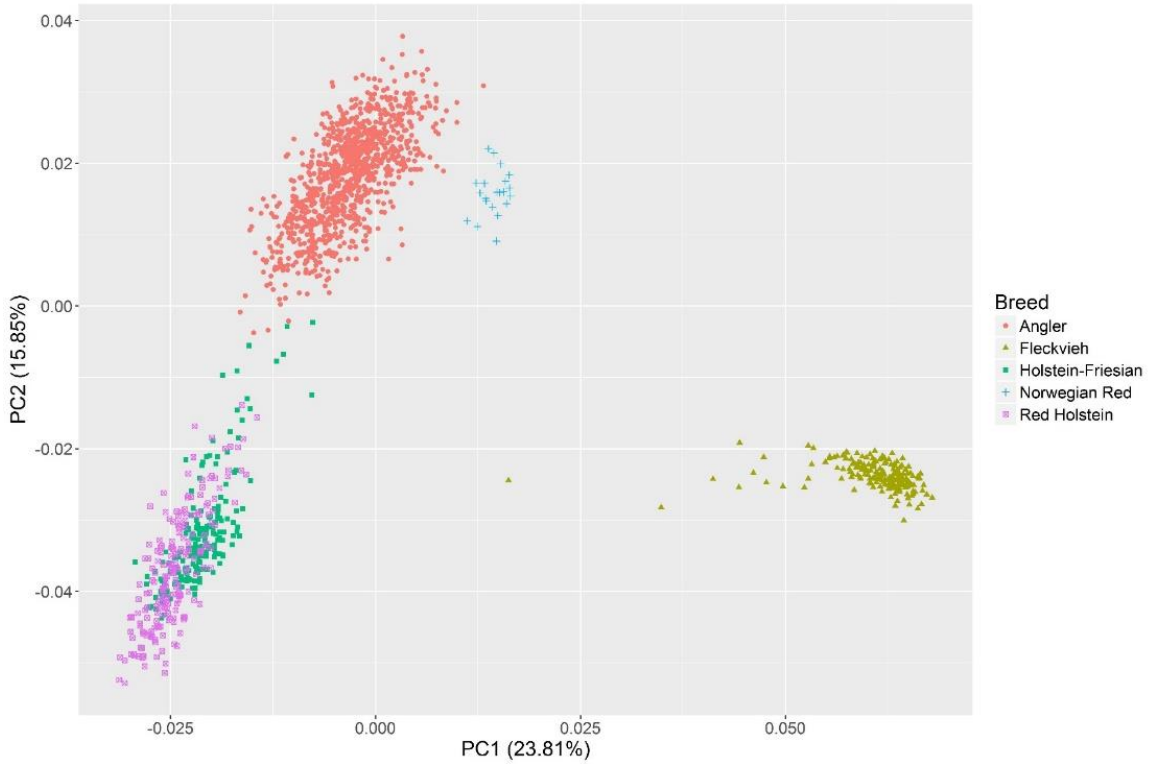



**Figure 4.1** Plot of the first two principal components (PCs) for the dataset of the simulated base population $G_0$. The analysis was based on 1,621 individuals and 23,448 SNPs. Different colors and shapes represent individuals from different breeds.

**Table 4.1** Average kinship ($f_{SEG}$) among five breeds of the simulated base generation $G_0$ based on shared segments.

| | **Angler** | **Red Holstein** | **Holstein-Friesian** | **Norwegian Red** | **Fleckvieh** |
|---|---|---|---|---|---|
| **Angler** | 0.048 | 0.039 | 0.037 | 0.017 | 0.004 |
| **Red Holstein** | | 0.110 | 0.085 | 0.007 | 0.004 |
| **Holstein-Friesian** | | | 0.095 | 0.007 | 0.004 |
| **Norwegian Red** | | | | 0.086 | 0.002 |
| **Fleckvieh** | | | | | 0.061 |

**Table 4.2** Basic statistics of the simulated true breeding values of each breed group of base generation $G_0$.

| Breed | N | Mean | SD |
|---|---|---|---|
| Angler | 1000 | 0.560 | 0.444 |
| Fleckvieh | 200 | 0.367 | 0.275 |
| Holstein-Friesian | 200 | 2.160 | 0.275 |
| Norwegian Red | 200 | 2.390 | 0.270 |
| Red Holstein | 21 | 2.431 | 0.276 |

## 4.3.2 Values of each parameter obtained in five scenarios

The mean and standard deviations of the parameter values at the starting stage [base generation ($G_0$)] and final stage [10$^{th}$ generation ($G_{10}$)] for all scenarios are shown in Table 4.3. The mean and standard deviations were estimated from five replicates. The values of the corresponding parameter for all generations can be found in Tables 4.S2 – 4.S8 [See Additional file 1 Table 4.S2-8].

### 4.3.2.1 EBV and MC

Except for the reference scenario, the average EBV level of Angler cattle increased in varying degrees from generation to generation in all scenarios, which is shown in Figure 4.2 (left). The EBV level remained stable at ~0.558 in REF. The average increase of EBV was ~0.9 genic standard deviations per generation in TS and OCS-I. The average EBV in $G_{10}$ was very similar in both scenario TS and OCS-I. The genetic gain was lower in scenario OCS-II, which achieved an average EBV of 2.757 in $G_{10}$, and was considerably lower in scenario OCS-III, for which the mean EBV was 1.825 in $G_{10}$.

MC decreased slightly from 0.622 to 0.587 in the reference scenario because old introgressed haplotype segments were split by crossing over into smaller pieces and could no longer be detected in $G_{10}$. In contrast, with the increase of EBV in scenario TS and OCS-I, the level of MC increased to a different extent (Figure 4.2, right). In scenario TS, MC increased from 0.622 in $G_0$ to 0.676 in $G_4$ and became stable in later generations. Similarly, in scenario OCS-I, MC increased to 0.647 at $G_4$ and became stable afterward. For scenario OCS-II and OCS-III, MC was set as a constraint. Thus, the average MC values obtained in each generation were approximately equal to the threshold setting in the corresponding generation with a rather small

standard deviation; that is, estimated MC remained 0.618 in OCS-II and decreased 3% each generation in scenario OCS-III.

#### 4.3.2.2 Kinship $f_{SEG}$ and kinship $f_{SEG|N}$

Kinship $f_{SEG}$ and kinship $f_{SEG|N}$ increased from generation to generation to varying extents, except for scenario REF, which can be seen in Figure 4.3 (Left: $f_{SEG}$; Right: $f_{SEG|N}$). Kinship $f_{SEG}$ had a small reduction in REF from 0.048 to 0.044, which was because old segments were split into smaller pieces, so after some generations, the pieces were no longer involved in shared segments. Kinship increased the most in scenario TS, which moved from 0.048 in $G_0$ to 0.115 in $G_{10}$. Kinship $f_{SEG}$ was set as a constraint in the other three scenarios. For scenario OCS-I, the $f_{SEG}$ value of each generation equals to the corresponding value of the constraint setting. For scenario OCS-II, in generation $G_{10}$, the $f_{SEG}$ value increased to 0.085, which is lower than the constraint setting. The smallest mean kinship (0.073) was obtained for scenario OCS-III.

Estimated kinship $f_{SEG|N}$ decreased from 0.061 in $G_0$ to 0.048 in $G_{10}$ for the reference scenario because some old introgressed segments were split into small pieces by crossing over, so the alleles included in the segments were classified as native and contributed to the estimated diversity at native alleles. Kinship $f_{SEG|N}$ increased faster in scenarios TS and OCS-I than kinship $f_{SEG}$. The value increased from 0.061 in $G_0$ to 0.157 in TS and to 0.136 in OCS-I. For scenarios OCS-II and OCS-III, $f_{SEG|N}$ was set as a constraint parameter. In all generations of both scenarios, the $f_{SEG|N}$ values were equal to the corresponding constraint setting of ub. $f_{SEG|N}$, with a standard deviation close to zero.

**Table 4.3** Basic statistics of each parameter achieved in base generation $G_0$ and $G_{10}$ for each selection scenario.

| | **Parameters**[a] | | | | | | |
|---|---|---|---|---|---|---|---|
| | **EBV** | **MC** | $\mathbf{f_{SEG}}$ | $\mathbf{f_{SEG|N}}$ | $\mathbf{H_O}$ | $\boldsymbol{\sigma^2_{TBV}}$ | $\boldsymbol{\sigma^2_A}$ |
| **Beginning of selection** | | | | | | | |
| $G_0$ | 0.561±0.000 | 0.622±0.000 | 0.048±0.000 | 0.061±0.000 | 0.367±0.000 | 0.197±0.000 | 0.075±0.000 |
| **End of selection ($G_{10}$)** | | | | | | | |
| *REF* | 0.558±0.020 | 0.587±0.002 | 0.044±0.001 | 0.048±0.001 | 0.364±0.001 | 0.091±0.005 | 0.075±0.001 |
| *TS* | 3.002±0.062 | 0.679±0.005 | 0.115±0.005 | 0.157±0.009 | 0.346±0.002 | 0.044±0.002 | 0.049±0.001 |
| *OCS-I* | 2.915±0.026 | 0.638±0.008 | 0.094±0.001 | 0.136±0.004 | 0.351±0.001 | 0.049±0.002 | 0.052±0.002 |
| *OCS-II* | 2.757±0.059 | 0.617±0.001 | 0.085±0.001 | 0.104±0.001 | 0.353±0.001 | 0.056±0.001 | 0.054±0.002 |
| *OCS-III* | 1.825±0.106 | 0.455±0.001 | 0.073±0.001 | 0.104±0.001 | 0.355±0.001 | 0.065±0.003 | 0.063±0.002 |

EBV, estimated breeding value; MC, migrant contribution; $f_{SEG}$, kinship; $f_{SEG|N}$, kinship at native alleles; $H_O$, average heterozygosity; $\sigma^2_{TBV}$, variance of true breeding value; $\sigma^2_A$, genic variance.

[a]Parameters estimated in each generation of each scenario.

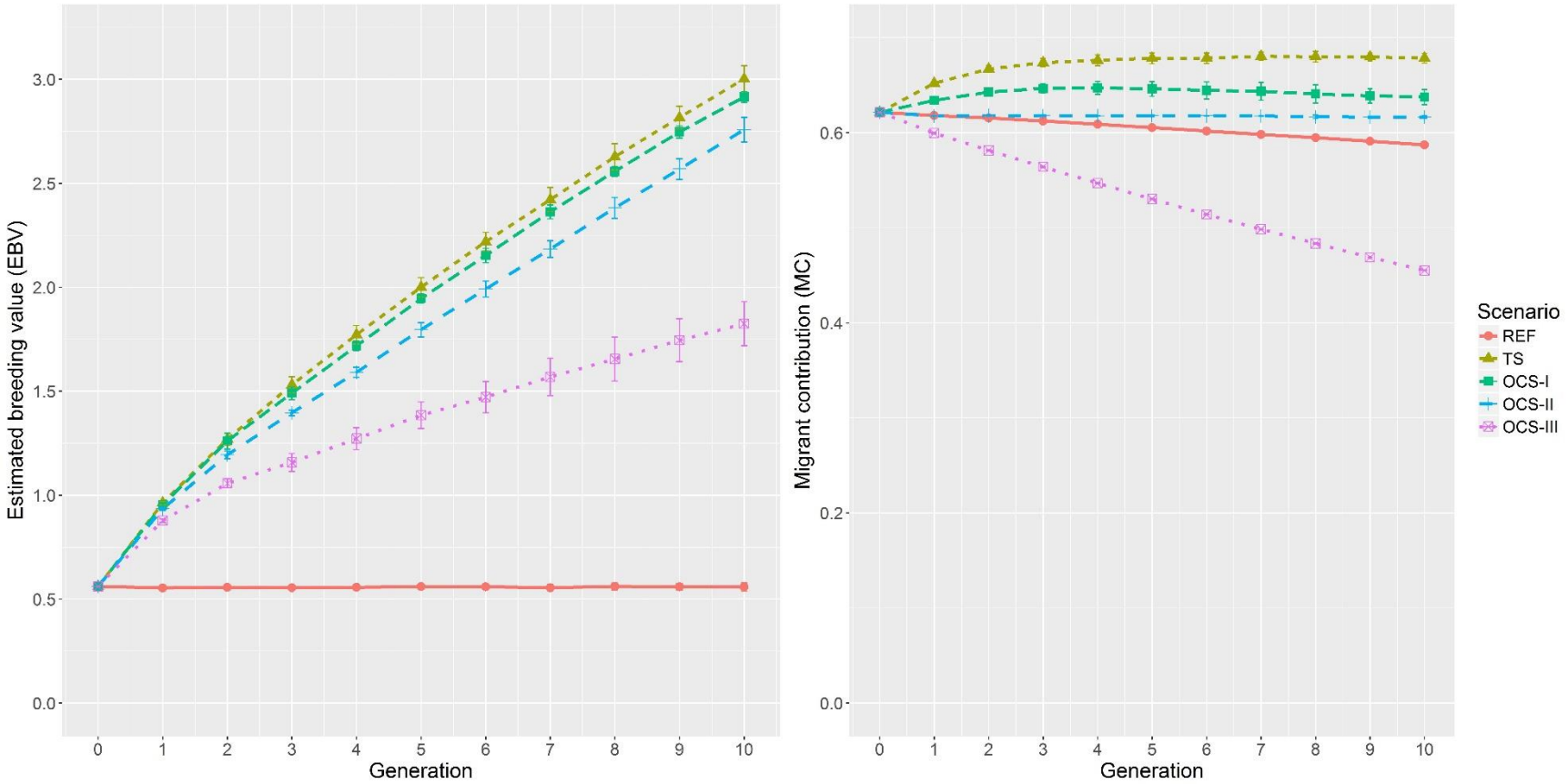



**Figure 4.2** Average estimated breeding values (Left) and migrant contribution (Right) achieved in each generation of each selection scenario. $f_{SEG}$, kinship; $f_{SEG|N}$, kinship at native alleles; EBV, estimated breeding value; MC, migrant contribution; OCS, optimum contribution selection; OCS-I, traditional OCS method; OCS-II, OCS with constraint on kinship $f_{SEG}$, kinship at native alleles $f_{SEG|N}$ and MC; OCS-III, OCS with constraint on kinship $f_{SEG}$, kinship at native alleles $f_{SEG|N}$ and reduced level of MC; REF, reference scenario; TS, truncation selection.

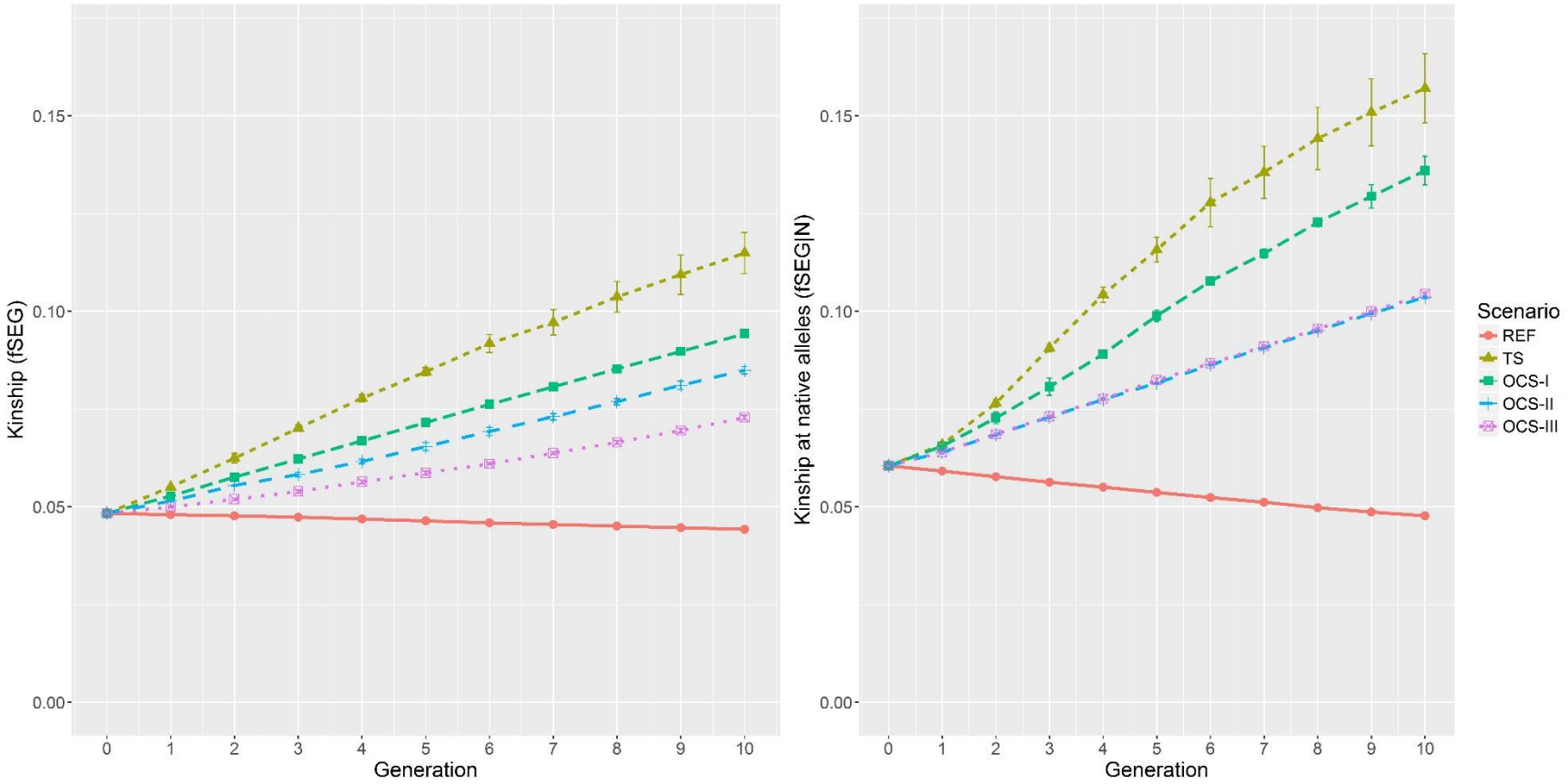



**Figure 4.3** Average kinship $f_{SEG}$ (Left) and kinship at native alleles $f_{SEG|N}$ (Right) achieved in each generation of each selection scenario. $f_{SEG}$, kinship; $f_{SEG|N}$, kinship at native alleles; EBV, estimated breeding value; MC, migrant contribution; OCS, optimum contribution selection; OCS-I, traditional OCS method; OCS-II, OCS with constraint on kinship $f_{SEG}$, kinship at native alleles $f_{SEG|N}$ and MC; OCS-III, OCS with constraint on kinship $f_{SEG}$, kinship at native alleles $f_{SEG|N}$ and reduced level of MC; REF, reference scenario; TS, truncation selection.

#### 4.3.2.3 Diversity parameters ($H_O$, $\sigma^2_{TBV}$ and $\sigma^2_A$)

The diversity parameters are shown in Figure 4.4 (Left: $H_O$; Middle: $\sigma^2_{TBV}$; Right: $\sigma^2_A$). As expected, all diversity values in $G_{10}$ of REF are higher than the corresponding values of all the other scenarios. The value of $H_O$ and $\sigma^2_A$ remained nearly unchanged from $G_0$ to $G_{10}$. The value of $\sigma^2_{TBV}$ decreased considerably from 0.197 to 0.091 from $G_0$ to $G_{10}$, which is still higher than the level of all the other scenarios (see Table 4.3). $\sigma^2_{TBV}$ was larger than $\sigma^2_A$ in $G_0$, which was due to the effects caused by different chromosomes being correlated.

For all scenarios, $H_O$ is relatively stable compared to $\sigma^2_{TBV}$ and $\sigma^2_A$. The greatest reduction of $H_O$ was found in scenario TS, which moved from 0.367 in $G_0$ to 0.346 in $G_{10}$. A similar trend but a faster reduction showed the genic variance $\sigma^2_A$. In scenario TS, the average $\sigma^2_A$ decreased from 0.075 in $G_0$ to 0.049 in $G_{10}$. A higher $\sigma^2_A$ value in $G_{10}$ was achieved in OCS-I (0.052) and OCS-II (0.054), and the highest genic variance was maintained in OCS-III (0.063).

In all scenarios, the level of $\sigma^2_{TBV}$ decreased considerably from $G_0$ to $G_1$. Thereafter, it decreased at a slower rate and approached the level of the genic variance around $G_6$. For the scenarios with selection, the variance of true breeding values in $G_{10}$ was very similar to the genic variance. It was 0.044 in TS, 0.049 in OCS-I, 0.056 in OCS-II, and the highest level was maintained in scenario OCS-III (0.065). In all scenarios, the average $\sigma^2_{TBV}$ level was much higher than the average genic variance $\sigma^2_A$ in the first generations. This can be seen in Figure 4.S3 [See Additional file 4 Figure 4.S3]) at the example of scenario OCS-II.

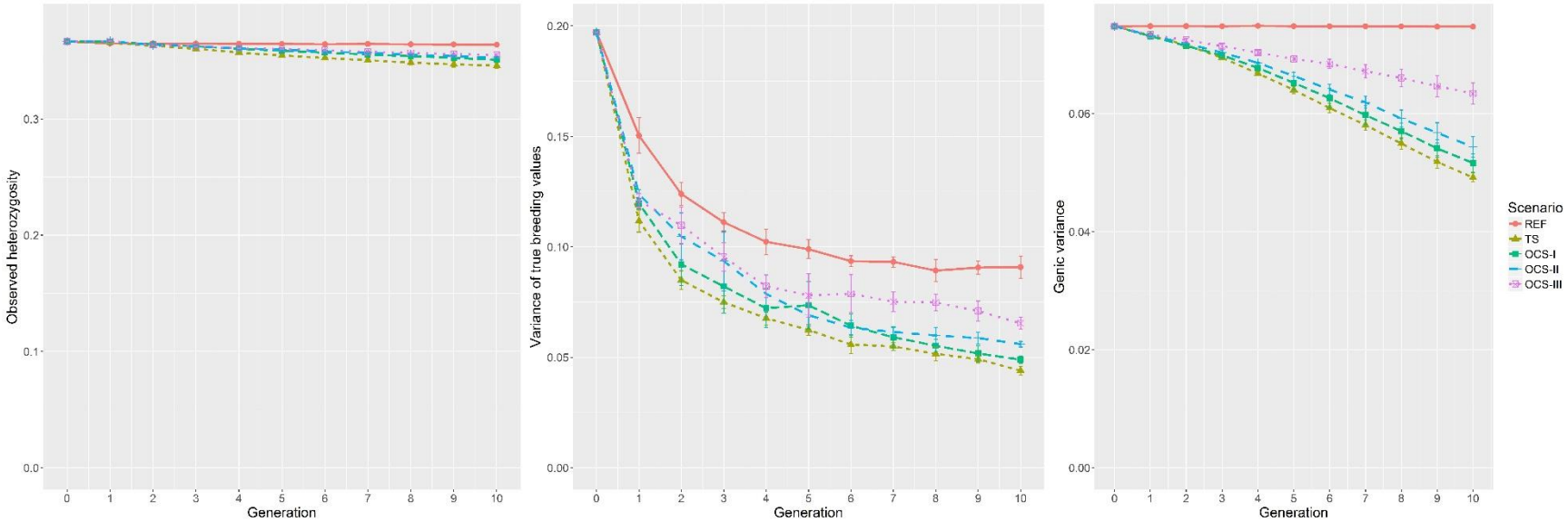



**Figure 4.4** Average observed heterozygosity $H_O$ (Left), variance of true breeding values $\sigma^2_{TBV}$ (Middle) and genetic variance $\sigma^2_A$ (Right) achieved in each generation of each selection scenario.

$f_{SEG}$, kinship; $f_{SEG|N}$, kinship at native alleles; EBV, estimated breeding value; MC, migrant contribution; OCS, optimum contribution selection; OCS-I, traditional OCS method; OCS-II, OCS with constraint on kinship $f_{SEG}$, kinship at native alleles $f_{SEG|N}$ and MC; OCS-III, OCS with constraint on kinship $f_{SEG}$, kinship at native alleles $f_{SEG|N}$ and reduced level of MC; REF, reference scenario; TS, truncation selection.

# 4.4 Discussion

In this study, we evaluated the long-term performance of five different scenarios for maximizing genetic gain in the context of conserving breeds with historic introgression using the example of Angler cattle. Migrant contributions and kinships at native alleles based on shared haplotype segments were restricted in some scenarios. A large proportion of the Angler breed's genetic background was contributed by other breeds, especially Holstein, which is in accordance with results obtained from pedigree records (Wang *et al.* 2017). Truncation selection achieved the highest genetic gain among the five scenarios with the highest degree of reduction in genetic diversity. Traditional OCS (OCS-I) achieved a slightly lower genetic gain and a slightly higher genetic diversity compared to truncation selection. However, both are not appropriate for the situation of Angler cattle, as both reduced the diversity at native alleles considerably and increased the MC. Constraining MC and kinship at native alleles enabled recovery of the genetic originality but also slowed the genetic progress in performance traits compared to truncation selection and traditional OCS.

## 4.4.1 Genetic progress vs. genetic conservation

Due to the protocol for simulating QTL effects, a positive correlation between MC and EBV was observed, which is in agreement with results obtained from the pedigree information (Wang *et al.* 2017). The positive correlation persisted in all generations in all five scenarios (data not shown). There was no genetic progress in REF due to the absence of selection. Truncation selection and traditional OCS achieved similar genetic gain. When MC and kinship at native alleles were constrained, the genetic gain in performance traits was reduced. Hence, to achieve maximum genetic gain, it is essential to allow for the introgression of foreign genetic material.

Maximizing genetic gain is not the only objective of a breeding program. To recover the genetic background of the original endangered population from admixtures, two goals must be set: maintain the genetic diversity at native alleles and remove the introgressed genomic material in the long run. The average MC of the population can be treated as a parameter for measuring genetic uniqueness. Among the five scenarios, truncation selection has the least ability to maintain genetic uniqueness. Although the situation improves in traditional OCS, the estimated MC level in $G_{10}$ is still higher than at the starting stage and higher than in the reference scenario

without selection. This is in accordance with what we obtained from OCS based on pedigrees. The reason why traditional OCS did not cause a larger increase in MC is probably that most migrant contributions were from related Holstein cattle. Thus, increasing genetic gain by increasing migrant contributions would also increase the average kinship in Angler cattle, so restricting average kinship implicitly restricted the migrant contribution. Similarly, in Amador *et al.* (2011), traditional OCS did not eliminate any exogenous representation but kept the value constant, irrespective of the number of generations elapsed before management started. Genetic originality could be maintained with the OCS-II method, while genetic gain was only marginally lower, as in traditional OCS and truncation selection.

In OCS-III, the EBV level kept increasing throughout all generations, even though the original genetic background was gradually reconstructed and the highest diversities were maintained with this method. Compared to OCS-II, the reduced genetic progress in OCS-III is directly linked to the strictness of the constraint MC setting. Due to the conflict between achieving genetic gain and maintaining genetic uniqueness, a breeding organization should choose MC constraint settings carefully to achieve both breeding purposes.

In truncation selection, 26 sires with the highest breeding values were selected along with 500 dams to achieve an effective population size of 100 in each generation. However, the formula from which the number of selected sires was obtained did not take into account that the individuals with the highest breeding values were related because they had high genetic contributions from closely related Holstein ancestors. Thus, the rate of inbreeding in truncation selection was higher than in the OCS scenarios. Compared to truncation selection, traditional OCS has good performance in controlling inbreeding via restricting average relatedness in the offspring.

### 4.4.2 Different kinship estimators

The predictions from $f_{SEG}$ and $f_{SEG|N}$ from *optiSel* were close to the values estimated from offspring haplotypes (results not shown). They were, however, slightly larger because some segments were split by crossing over into small pieces, so they did not contribute to the kinship estimated from offspring haplotypes. This indicates that the estimate obtained from offspring haplotypes is slightly biased. The rate of inbreeding estimated from segments remained 0.5% per generation in traditional OCS, in accordance with the constraint level setting.

Kinship $f_{SEG}$ estimates the probability that randomly chosen alleles are IBD. However, it lacks the ability to distinguish whether the alleles originated from native or migrant ancestors. In scenario OCS-I, where $f_{SEG}$ was restricted, the increasing rate of $f_{SEG|N}$ was higher than the increasing rate of $f_{SEG}$. This suggests that restricting only $f_{SEG}$ had the consequence that diversity at introgressed segments was maintained, which tend to have higher breeding values. But a depletion of diversity at native segments could not be avoided. Because kinship and kinship at native alleles are correlated, restricting $f_{SEG|N}$ implicitly restricted $f_{SEG}$, so in scenarios OCS-II and OCS-III, the mean kinship $f_{SEG}$ was lower than the corresponding constraint setting. This suggests that the constraint for $f_{SEG}$ could be skipped if $f_{SEG|N}$ and MC are constrained. Similar results were obtained from pedigree information by Wang *et al.* (2017).

### 4.4.3 Migrant contributions

In general, it must be distinguished whether migrant contributions predominantly originate from closely related ancestors originating from a single high-yielding breed, or if different unrelated breeds have been used for upgrading. In the Angler breed, they predominantly originated from related Holstein ancestors, so reducing MC in OCS-III was meant to reduce the amount of genetic material contributed by Holstein cattle, which had a positive effect on the genetic diversity. Thus, the mean kinship $f_{SEG}$ in OCS-III was smaller than in all other scenarios.

### 4.4.4 Criteria for detecting shared segments

It has been suggested that the marker density of the SNP chip used, the minimum length of the shared segment, the number of genotyping errors allowed, and the minimum number of SNPs allowed in a single shared segment are likely to remarkably influence kinship estimates based on shared segments (Peripolli *et al.* 2016). However, to date, there is a lack of consensus in establishing the criteria for determining these parameters, which makes it difficult to compare results from different studies. In this paper, the minimum number of markers in a segment was 20. A shorter minimum length for shared segments allows detection of more ancient inbreeding from common ancestors occurring many generations back (Curik *et al.* 2014), but it also increases the probability that segments that are identical by chance are considered to be IBD. The minimum length of a segment was 2.50 Mb because, in this case, the correlation between the contribution from the Holstein breed estimated from the pedigree and genotype was high (0.93, data not shown), and the genetic contribution from the Fleckvieh breed was low (~0.02), in accordance with pedigree records. The average migrant contribution of the Angler population was 0.62, which is also similar to the average migrant contribution level obtained from the pedigree (Wang *et al.* 2017). If shorter segments were also to be used, then kinships of individuals would be affected more by very old common ancestors and would consequently be higher.

### 4.4.5 Reduction of estimates in unselected populations

Estimated parameter values for MC, $f_{SEG}$ and $f_{SEG|N}$ decreased slightly from generation to generation in REF, even though there was no selection in this scenario. This reduction of the above three parameters was caused by recombination, which shortened the length of the haplotype segments (Stam 1980) until they became too short to meet the criteria for being segments, which led to the reduction of $f_{SEG}$. Moreover, if recombination occurred near a particular marker position at an introgressed haplotype segment, then the segment containing the marker could no longer be detected in other breeds. Hence, the marker failed to meet the criteria of belonging to a foreign segment. This gave rise to the reduction of estimated MC. Moreover, since the marker was now classified as native, it contributed to the diversity at native alleles, which caused a reduction in $f_{SEG|N}$. Consequently, the estimated MC should be compared with the estimates obtained from the reference scenario rather than with generation $G_0$. In particular, in scenario OCS-II, in which the constraint for MC was set equal to the MC

in generation $G_0$, only the estimate of MC was kept constant, whereas the true MC was effectively increased. There are two possibilities to avoid this increase. Either the constraint for MC in generation $G_t$ is set equal to the mean MC in generation $G_t$ (rather than the mean MC in generation $G_0$), or another method could be used to estimate the origins of the haplotype segments in generation $G_{t+1}$. That is, the origin of a marker could be set equal to the origin of the marker in the parental haplotype from which it originates.

### 4.4.6 Genetic diversity parameters

Different parameters can be used to measure genetic diversity, such as the percentage of polymorphic loci, the number of alleles per locus, expected heterozygosity, etc. (Harper and Hawksworth 1994). The genetic variation within a breed is of major importance for conservation of local breeds. In addition to $f_{SEG}$ and $f_{SEG|N}$, three further parameters were considered for evaluating the level of genetic diversity, i.e. the average observed heterozygosity ($H_O$), the variance of true breeding value ($\sigma^2_{TBV}$) and the genic variance ($\sigma^2_A$). Restricting kinship at native alleles and migrant contributions not only had an impact on recovering the original genetic background but also showed the most potential in conserving genetic diversity among all scenarios. In this study, a similar decreasing pattern of $H_O$ and $\sigma^2_A$ was observed, with a smaller extent of reduction for parameter $H_O$. This is because $H_O$ is predominantly influenced by neutral alleles (Gregorius 1978).

The additive genetic variance $\sigma^2_{TBV}$ was substantially larger than the genic variance in the first generations and decreased to a large amount from generation $G_0$ to generation $G_5$. This was predominantly because the genetic effects of different chromosomes were correlated in the Angler breed. The contribution of the covariance between different chromosomes to the variance of TBV was 0.089, so in the absence of the covariance, the variance of TBV should be 0.108. The Angler cattle in generation $G_0$ had different contributions from the high-yielding Holstein cattle. For an individual with a high contribution from Holstein cattle, the breeding values of all chromosomes tended to be high, whereas for an individual with a low contribution from Holstein, the breeding values of all chromosomes tended to be low. Consequently, in the first generations, there was covariance between effects of different chromosomes, which contributed to the variance of the breeding values. Additionally, the Bulmer Effect (Bulmer 1971) and the changes in linkage disequilibrium due to selection (Bijma 2012; Gorjanc *et al.* 2015) contributed to the difference between $\sigma^2_{TBV}$ and $\sigma^2_A$ .

## 4.5 Conclusions

Advanced OCS strategies enable achieving a balance between the different breeding goals of populations with historic introgression, which are to improve the genetic progress, to recover the original genetic background and to conserve genetic diversity. Truncation selection and traditional OCS achieved the highest genetic gain, but both reduced the genetic originality of the breed by depleting diversity at native alleles and increasing migrant contributions. Maintaining genetic originality, however, is crucial for conserving breeds with historical introgression. The inclusion of MC and kinship at native alleles as additional constraints in OCS showed great potential for conservation. Recovering the original genetic background is possible but requires many generations of selection and reduces the genetic progress. Thus, it is essential to set an appropriate constraint for MC in order to balance both breeding goals, which are to achieve genetic progress and to recover the original genetic background of local breeds.

### Acknowledgements

We thank the editor and two anonymous reviewers for their comments and advice. The study was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG). The authors declare that they have no competing interests.

# References

Alexander D. H., Novembre J., Lange K., 2009 Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19: 1655–1664.

Amador C., Toro M. Á., Fernández J., 2011 Removing exogenous information using pedigree data. Conserv. Genet. 12: 1565–1573.

Amador C., Fernández J., Meuwissen T. H. E., 2013 Advantages of using molecular coancestry in the removal of introgressed genetic material. Genet. Sel. Evol. 45: 13.

Baran Y., Pasaniuc B., Sankararaman S., Torgerson D. G., Gignoux C., *et al.*, 2012 Fast and accurate inference of local ancestry in Latino populations. Bioinformatics 28: 1359–1367.

Bennewitz J., Meuwissen T. H. E., 2005 Estimation of extinction probabilities of five German cattle breeds by population viability analysis. J. Dairy Sci. 88: 2949–2961.

Bijma P., 2012 Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. J. Anim. Breed. Genet. 129: 345–358.

Browning S. R., Browning B. L., 2007 Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81: 1084–1097.

Bulmer M. G., 1971 The effect of selection on genetic variability. Am. Nat. 105: 201–211.

Caballero A., Toro M. Á., 2002 Analysis of genetic diversity for the management of conserved subdivided populations. Conserv. Genet. 3: 289–299.

Chang C. C., Chow C. C., Tellier L. C., Vattikuti S., Purcell S. M., *et al.*, 2015 Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4: 7.

Curik I., Ferenčaković M., Sölkner J., 2014 Inbreeding and runs of homozygosity: A possible solution to an old problem. Livest. Sci. 166: 26–34.

de Cara M. Á. R., Fernández J., Toro M. Á., Villanueva B., 2011 Using genome-wide information to minimize the loss of diversity in conservation programmes. J. Anim. Breed. Genet. 128: 456–464.

de Cara M. Á. R., Villanueva B., Toro M. Á., Fernández J., 2013 Using genomic tools to maintain diversity and fitness in conservation programmes. Mol. Ecol. 22: 6091–6099.

Eding H., Meuwissen T. H. E., 2001 Marker-based estimates of between and within population kinships for the conservation of genetic diversity. J. Anim. Breed. Genet. 118: 141–159.

Eynard S. E., Windig J. J., Hiemstra S. J., Calus M. P. L., 2016 Whole-genome sequence data uncover loss of genetic diversity due to selection. Genet. Sel. Evol. 48: 33.

Falconer D. S., Mackay T. F., 1996 *Introduction to quantitative genetics.* Harlow: Pearson Ediucation Limited.

Frkonja A., Gredler B., Schnyder U., Curik I., Sölkner J., 2012 Prediction of breed composition in an admixed cattle population. Anim. Genet. 43: 696–703.

Gómez-Romano F., Villanueva B., Sölkner J., Cara M. Á. R. de, Mészáros G., *et al.*, 2016 The use of coancestry based on shared segments for maintaining genetic diversity. J. Anim. Breed. Genet. 133: 357–365.

Gorjanc G., Bijma P., Hickey J. M., 2015 Reliability of pedigree-based and genomic evaluations in selected populations. Genet. Sel. Evol. 47: 65.

Gregorius H.-R., 1978 The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance. Math. Biosci. 41: 253–271.

Grundy B., Villanueva B., Woolliams J. A., 1998 Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. Genet. Res. (Camb). 72: 159–168.

Gusev A., Lowe J. K., Stoffel M., Daly M. J., Altshuler D., *et al.*, 2009 Whole population, genome-wide mapping of hidden relatedness. Genome Res. 19: 318–26.

Harper J. L., Hawksworth D. L., 1994 Biodiversity: measurement and estimation. Phisophical Trans. R. Soc. London 345: 5–12.

Keller M. C., Visscher P. M., Goddard M. E., 2011 Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics 189: 237–249.

McTavish E. J., Hillis D. M., 2015 How do SNP ascertainment schemes and population demographics affect inferences about population history? BMC Genomics 16: 266.

Mészáros G., Boison S. A., Pérez O'Brien A. M., Ferenčaković M., Curik I., *et al.*, 2015 Genomic analysis for managing small and endangered populations: A case study in Tyrol Grey cattle. Front. Genet. 6: 1–12.

Meuwissen T. H. E., 1997 Maximizing the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75: 2575–2583.

Meuwissen T. H. E., Hayes B. J., Goddard M. E., 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829.

Meuwissen T. H. E., 2009 Genetic management of small populations: A review. Acta Agr Scand A-AN 59: 71–79.

Nielsen R., 2000 Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics 154: 931–942.

Peripolli E., Munari D. P., Silva M. V. B., Lima A. L. F., Irgang R., *et al.*, 2016 Runs of homozygosity: current knowledge and applications in livestock. Anim. Genet. 48: 255–271.

Pfaff B., 2014 The R package *cccp*: Design for solving cone constrained convex programs. R Financ.: http://www.pfaffikus.de/files/conf/rif/rif2014.pdf.

Price A. L., Patterson N. J., Plenge R. M., Weinblatt M. E., Shadick N. A., *et al.*, 2006 Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38: 904–909.

Pritchard J. K., Stephens M., Donnelly P., 2000 Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Rodríguez-Ramilo S. T., García-Cortés L. A., Cara M. Á. R. de, 2015 Artificial selection with traditional or genomic relationships: consequences in coancestry and genetic diversity. Front. Genet. 6: 127.

Ron M., Blanc Y., Band M., Ezra E., Weller J. I., 1996 Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement. J. Dairy Sci. 79: 676–681.

Stam P., 1980 The distribution of the fraction of the genome identical by descent in finite random mating populations. Genet. Res. 35: 131.

Toro M. Á., Villanueva B., Fernández J., 2014 Genomics applied to management strategies in conservation programmes. Livest. Sci. 166: 48–53.

VanRaden P. M., 2008 Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414–23.

Visscher P. M., Medland S. E., Ferreira M. A. R., Morley K. I., Zhu G., *et al.*, 2006 Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet. 2: 0316–0325.

Wang Y., Bennewitz J., Wellmann R., 2017 Novel optimum contribution selection methods accounting for conflicting objectives in breeding programs for livestock breeds with historical migration. Genet. Sel. Evol. 49: 45.

Wellmann R., Hartwig S., Bennewitz J., 2012 Optimum contribution selection for conserved populations with historic migration. Genet. Sel. Evol. 44: 34.

Wellmann R., 2017 *optiSel*: Optimum Contribution Selection and Population Genetics. R Packag. version 0.9.1. https://cran.r-project.org/web/packages/optiSel/optiSel.pdf.

Weng Z.-Q., Saatchi M., Schnabel R. D., Taylor J. F., Garrick D. J., 2014 Recombination locations and rates in beef cattle assessed from parent-offspring pairs. Genet. Sel. Evol. 46: 34.

Woolliams J. A., Berg P., Dagnachew B. S., Meuwissen T. H. E., 2015 Genetic contributions and their optimization. J. Anim. Breed. Genet. 132: 89–99.

Yang J., Benyamin B., McEvoy B. P., Gordon S., Henders A. K., et al., 2010 Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42: 565–569.

# Appendix

## Native segments

Each individual i has a maternal haplotype $H_{1i}$ and a paternal haplotype $H_{2i}$. For haplotype $H_{vi}$ of Angler i, the frequency of the segment at marker m in the set of animals from reference breed B was calculated as:

$$p_{vi}(m;B) = \frac{1}{2 * N_B} \sum_{j \in B} 1_{m \in S(H_{vi}, H_{1j})} + 1_{m \in S(H_{vi}, H_{2j})},$$

where $N_B$ is the number of individuals from breed B, set $S(H_{vi}, H_{wj})$ contains all markers belonging to segments, which are identical in haplotypes $H_{vi}$ and $H_{wj}$, and $1_{m \in S(H_{vi}, H_{2j})} = 1$ if marker m belongs to such a segment.

A marker m from haplotype $H_{vi}$ was classified as native ($N_{vi}(m) = 1$) if the frequency of the segment containing marker m is smaller than $\varepsilon = 0.01$ in all reference breeds. That is,

$$N_{vi}(m) = 1 \Leftrightarrow \max_B p_{vi}(m;B) < \varepsilon,$$

where the maximum was taken over all reference breeds, which were Fleckvieh, Holstein-Friesian, Red Holstein, and Norwegian Red in our study.

The native contribution $N_i$ of individual i is the proportion of the genome included in native segments. That is,

$$N_i = \frac{1}{2L} \sum_m L_m \cdot (N_{1i}(m) + N_{2i}(m)).$$

where $L_m$ is the length of the genome region in Mb represented by marker m, and L is the length of the genome in Mb. The migrant contribution of individual i is $MC_i = 1 - N_i$.

## Kinship $\mathbf{f_{SEG}}$ and $\mathbf{f_{SEG|N}}$

The kinship $f_{SEG}$ between individual i and j (element of matrix $\mathbf{f_{SEG}}$) is the probability that two alleles taken at random from individual i and j belong to identical segments:

$$f_{SEG_{ij}} = \frac{1}{4}\sum_{v,w=1}^{2}\frac{\sum_m L_m \cdot 1_{m\in S(H_{vi},H_{wj})}}{L},$$

where all parameters involved are the same as previously explained.

The kinship $f_{SEG|N}$ in the offspring is the conditional probability that two alleles taken at random from the offspring belong to identical segments, given that they are native. The mean kinship $f_{SEG|N}$ of the following generation is estimated as:

$$f_{SEG|N} = \frac{\mathbf{c'f_{SEG\&N}c}}{\mathbf{c'f_N c}},$$

where $\mathbf{f_{SEG\&N}}$ is a matrix containing the probabilities that two alleles taken at random from both individuals belong to identical segments and are native, and $\mathbf{f_N}$ is a matrix containing the probabilities that two alleles taken at random from both individuals are native:

$$f_{SEG\&N}(i,j) = \frac{1}{4}\sum_{v,w=1}^{2}\frac{\sum_m L_m \cdot N_{vi}(m)\cdot N_{wj}(m)\cdot 1_{m\in S(H_{vi},H_{wj})}}{L},$$

$$f_N(i,j) = \frac{1}{4}\sum_{v,w=1}^{2}\frac{\sum_m L_m \cdot N_{vi}(m)\cdot N_{wj}(m)}{L}.$$

# Additional files

**Additional file 1 Table 4.S1** Constraint settings of each parameter of each generation for specified scenarios

| | $\mathbf{ub.f_{SEG}}$ | $\mathbf{ub.f_{SEG|N}}$ | **ub. MC** |
|---|---|---|---|
| $G_1$ | 0.053 | 0.065 | 0.603 |
| $G_2$ | 0.058 | 0.070 | 0.585 |
| $G_3$ | 0.063 | 0.075 | 0.567 |
| $G_4$ | 0.067 | 0.079 | 0.550 |
| $G_5$ | 0.072 | 0.084 | 0.534 |
| $G_6$ | 0.077 | 0.088 | 0.518 |
| $G_7$ | 0.081 | 0.093 | 0.502 |
| $G_8$ | 0.086 | 0.097 | 0.487 |
| $G_9$ | 0.090 | 0.102 | 0.472 |
| $G_{10}$ | 0.095 | 0.106 | 0.458 |
| Applied in Scenario | *OCS-I, OCS-II, OCS-III* | *OCS-II, OCS-III* | *OCS-III* |

**Additional file 1 Table 4.S2** Descriptive statistics of estimated breeding values (EBV) achieved in each generation for each selection scenario

| **Scenario** | ***REF*** | | ***TS*** | | ***OCS-I*** | | ***OCS-II*** | | ***OCS-III*** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** |
| $G_0$ | 0.561 | 0.000 | 0.561 | 0.000 | 0.561 | 0.000 | 0.561 | 0.000 | 0.561 | 0.000 |
| $G_1$ | 0.554 | 0.014 | 0.963 | 0.008 | 0.954 | 0.008 | 0.935 | 0.007 | 0.878 | 0.009 |
| $G_2$ | 0.557 | 0.010 | 1.270 | 0.028 | 1.261 | 0.039 | 1.194 | 0.018 | 1.057 | 0.021 |
| $G_3$ | 0.555 | 0.010 | 1.531 | 0.039 | 1.491 | 0.031 | 1.396 | 0.014 | 1.158 | 0.043 |
| $G_4$ | 0.557 | 0.014 | 1.772 | 0.045 | 1.718 | 0.022 | 1.591 | 0.024 | 1.272 | 0.052 |
| $G_5$ | 0.560 | 0.012 | 2.000 | 0.046 | 1.946 | 0.022 | 1.796 | 0.034 | 1.385 | 0.064 |
| $G_6$ | 0.560 | 0.013 | 2.219 | 0.045 | 2.153 | 0.034 | 1.992 | 0.038 | 1.472 | 0.075 |
| $G_7$ | 0.555 | 0.016 | 2.421 | 0.058 | 2.363 | 0.034 | 2.183 | 0.040 | 1.569 | 0.090 |
| $G_8$ | 0.561 | 0.017 | 2.627 | 0.063 | 2.557 | 0.023 | 2.382 | 0.050 | 1.655 | 0.105 |
| $G_9$ | 0.559 | 0.017 | 2.815 | 0.055 | 2.747 | 0.030 | 2.568 | 0.049 | 1.745 | 0.103 |
| $G_{10}$ | 0.558 | 0.020 | 3.002 | 0.062 | 2.915 | 0.026 | 2.757 | 0.059 | 1.825 | 0.106 |

**Additional file 1 Table 4.S3** Descriptive statistics of migrant contribution (MC) achieved in each generation for each scenario

| **Scenario** | ***REF*** | | ***TS*** | | ***OCS-I*** | | ***OCS-II*** | | ***OCS-III*** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** |
| $G_0$ | 0.622 | 0.000 | 0.622 | 0.000 | 0.622 | 0.000 | 0.622 | 0.000 | 0.622 | 0.000 |
| $G_1$ | 0.618 | 0.001 | 0.652 | <0.001 | 0.634 | <0.001 | 0.618 | <0.001 | 0.600 | 0.001 |
| $G_2$ | 0.615 | 0.001 | 0.667 | 0.004 | 0.643 | 0.004 | 0.618 | 0.001 | 0.582 | <0.001 |
| $G_3$ | 0.612 | 0.001 | 0.674 | 0.004 | 0.647 | 0.005 | 0.618 | <0.001 | 0.564 | 0.001 |
| $G_4$ | 0.609 | 0.002 | 0.676 | 0.006 | 0.647 | 0.007 | 0.618 | <0.001 | 0.547 | 0.001 |
| $G_5$ | 0.605 | 0.002 | 0.678 | 0.006 | 0.646 | 0.008 | 0.618 | 0.001 | 0.530 | 0.001 |
| $G_6$ | 0.602 | 0.002 | 0.678 | 0.006 | 0.645 | 0.009 | 0.618 | <0.001 | 0.514 | <0.001 |
| $G_7$ | 0.598 | 0.002 | 0.681 | 0.004 | 0.644 | 0.009 | 0.618 | <0.001 | 0.499 | <0.001 |
| $G_8$ | 0.595 | 0.002 | 0.680 | 0.006 | 0.641 | 0.009 | 0.617 | 0.002 | 0.484 | <0.001 |
| $G_9$ | 0.591 | 0.002 | 0.680 | 0.004 | 0.639 | 0.008 | 0.617 | 0.001 | 0.469 | <0.001 |
| $G_{10}$ | 0.587 | 0.002 | 0.679 | 0.005 | 0.638 | 0.008 | 0.617 | 0.001 | 0.455 | <0.001 |

**Additional file 1 Table 4.S4** Descriptive statistics of kinship $f_{SEG}$ achieved in each generation for each scenario

| **Scenario** | ***REF*** | | ***TS*** | | ***OCS-I*** | | ***OCS-II*** | | ***OCS-III*** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** |
| $G_0$ | 0.048 | 0.000 | 0.048 | 0.000 | 0.048 | 0.000 | 0.048 | 0.000 | 0.048 | 0.000 |
| $G_1$ | 0.048 | <0.001 | 0.055 | <0.001 | 0.053 | <0.001 | 0.052 | <0.001 | 0.050 | <0.001 |
| $G_2$ | 0.048 | <0.001 | 0.063 | 0.001 | 0.058 | <0.001 | 0.056 | <0.001 | 0.052 | <0.001 |
| $G_3$ | 0.048 | <0.001 | 0.070 | 0.001 | 0.062 | <0.001 | 0.058 | 0.001 | 0.054 | <0.001 |
| $G_4$ | 0.047 | <0.001 | 0.078 | 0.001 | 0.067 | <0.001 | 0.062 | 0.001 | 0.056 | 0.001 |
| $G_5$ | 0.047 | <0.001 | 0.085 | 0.001 | 0.072 | <0.001 | 0.065 | 0.001 | 0.059 | 0.001 |
| $G_6$ | 0.046 | <0.001 | 0.092 | 0.002 | 0.076 | <0.001 | 0.069 | 0.001 | 0.061 | <0.001 |
| $G_7$ | 0.045 | <0.001 | 0.097 | 0.003 | 0.081 | <0.001 | 0.073 | 0.001 | 0.064 | <0.001 |
| $G_8$ | 0.045 | <0.001 | 0.104 | 0.004 | 0.085 | <0.001 | 0.077 | 0.001 | 0.067 | <0.001 |
| $G_9$ | 0.045 | <0.001 | 0.109 | 0.005 | 0.090 | <0.001 | 0.081 | 0.001 | 0.070 | 0.001 |
| $G_{10}$ | 0.044 | <0.001 | 0.115 | 0.005 | 0.094 | <0.001 | 0.085 | 0.001 | 0.073 | 0.001 |

**Additional file 1 Table 4.S5** Descriptive statistics of kinship at native alleles $f_{SEG|N}$ achieved in each generation for each scenario

| **Scenario** | ***REF*** | | ***TS*** | | ***OCS-I*** | | ***OCS-II*** | | ***OCS-III*** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** |
| $G_0$ | 0.061 | <0.000 | 0.061 | 0.000 | 0.061 | 0.000 | 0.061 | 0.000 | 0.061 | 0.000 |
| $G_1$ | 0.061 | <0.001 | 0.066 | <0.001 | 0.066 | <0.001 | 0.064 | <0.001 | 0.064 | <0.001 |
| $G_2$ | 0.059 | <0.001 | 0.077 | 0.001 | 0.073 | 0.002 | 0.069 | <0.001 | 0.069 | <0.001 |
| $G_3$ | 0.058 | <0.001 | 0.091 | 0.001 | 0.081 | 0.002 | 0.073 | <0.001 | 0.073 | <0.001 |
| $G_4$ | 0.056 | <0.001 | 0.104 | 0.002 | 0.089 | 0.001 | 0.078 | <0.001 | 0.078 | <0.001 |
| $G_5$ | 0.055 | <0.001 | 0.116 | 0.003 | 0.099 | 0.002 | 0.082 | <0.001 | 0.082 | 0.001 |
| $G_6$ | 0.054 | <0.001 | 0.128 | 0.006 | 0.108 | 0.001 | 0.086 | <0.001 | 0.087 | <0.001 |
| $G_7$ | 0.051 | <0.001 | 0.136 | 0.007 | 0.115 | 0.001 | 0.091 | <0.001 | 0.091 | <0.001 |
| $G_8$ | 0.050 | <0.001 | 0.144 | 0.008 | 0.123 | 0.001 | 0.095 | <0.001 | 0.096 | <0.001 |
| $G_9$ | 0.049 | 0.001 | 0.151 | 0.009 | 0.129 | 0.003 | 0.099 | <0.001 | 0.100 | <0.001 |
| $G_{10}$ | 0.048 | 0.001 | 0.157 | 0.009 | 0.136 | 0.004 | 0.104 | <0.001 | 0.104 | <0.001 |

**Additional file 1 Table 4.S6** Descriptive statistics of observed heterozygosity ($H_O$) achieved in each generation for each scenario

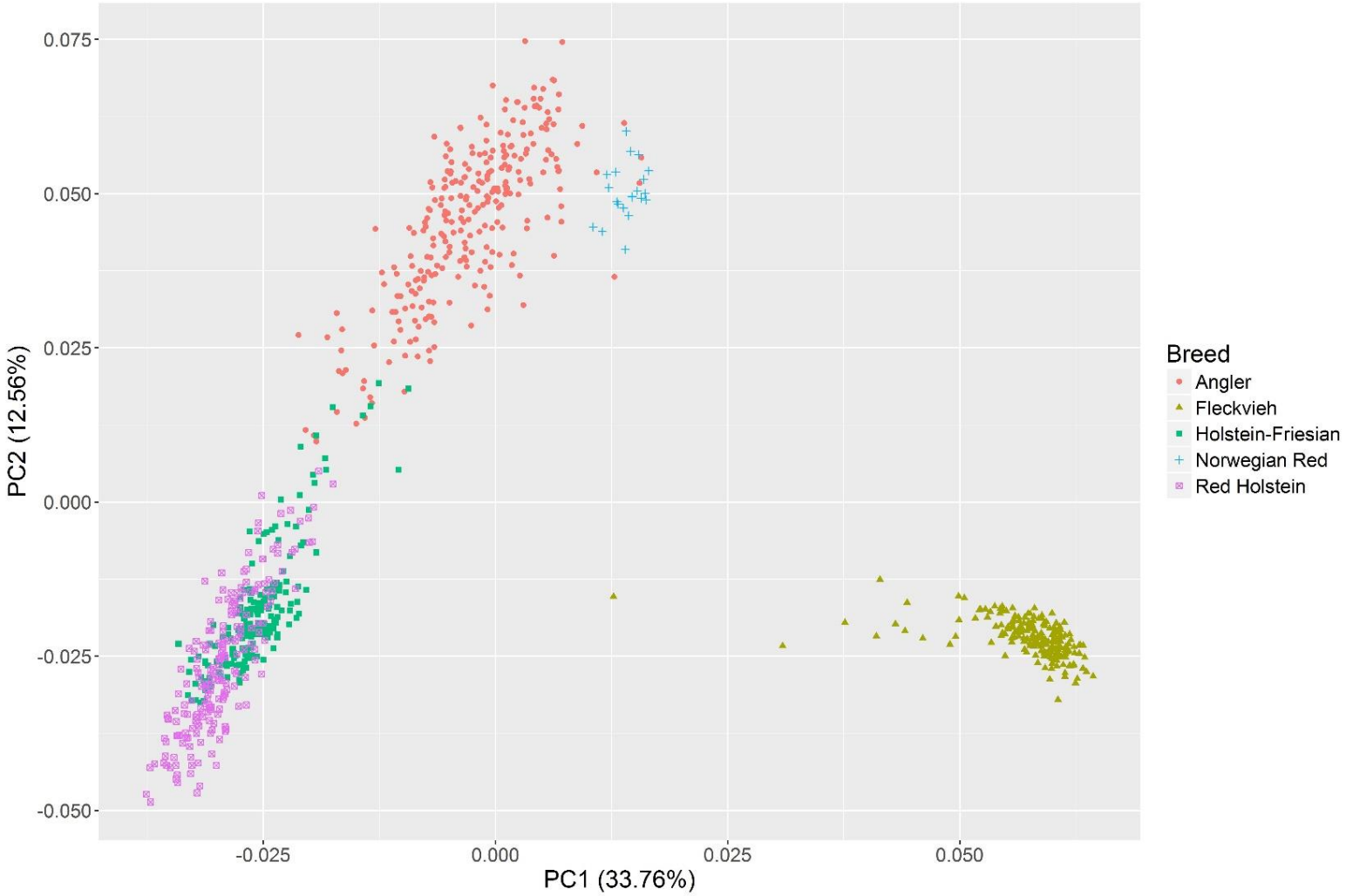
| **Scenario** | ***REF*** | | ***TS*** | | ***OCS-I*** | | ***OCS-II*** | | ***OCS-III*** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** |
| $G_0$ | 0.367 | 0.000 | 0.367 | 0.000 | 0.367 | 0.000 | 0.367 | 0.000 | 0.367 | 0.000 |
| $G_1$ | 0.367 | <0.001 | 0.365 | 0.001 | 0.367 | <0.001 | 0.367 | <0.001 | 0.367 | <0.001 |
| $G_2$ | 0.365 | <0.001 | 0.363 | <0.001 | 0.364 | 0.001 | 0.364 | <0.001 | 0.364 | 0.001 |
| $G_3$ | 0.365 | <0.001 | 0.360 | <0.001 | 0.362 | 0.001 | 0.363 | 0.001 | 0.362 | 0.001 |
| $G_4$ | 0.365 | <0.001 | 0.357 | 0.001 | 0.360 | 0.001 | 0.361 | 0.001 | 0.361 | <0.001 |
| $G_5$ | 0.365 | <0.001 | 0.355 | 0.001 | 0.359 | 0.001 | 0.360 | 0.001 | 0.360 | <0.001 |
| $G_6$ | 0.364 | <0.001 | 0.352 | 0.001 | 0.357 | 0.001 | 0.358 | <0.001 | 0.359 | <0.001 |
| $G_7$ | 0.364 | <0.001 | 0.351 | 0.002 | 0.356 | 0.001 | 0.357 | <0.001 | 0.358 | 0.001 |
| $G_8$ | 0.364 | <0.001 | 0.349 | 0.002 | 0.354 | 0.001 | 0.355 | 0.001 | 0.357 | 0.001 |
| $G_9$ | 0.364 | <0.001 | 0.347 | 0.003 | 0.352 | 0.001 | 0.354 | 0.001 | 0.356 | <0.001 |
| $G_{10}$ | 0.364 | <0.001 | 0.346 | 0.002 | 0.351 | 0.001 | 0.353 | 0.001 | 0.355 | 0.001 |

**Additional file 1 Table 4.S7** Descriptive statistics of the variance of true breeding values ($\sigma^2_{TBV}$) achieved in each generation for each scenario

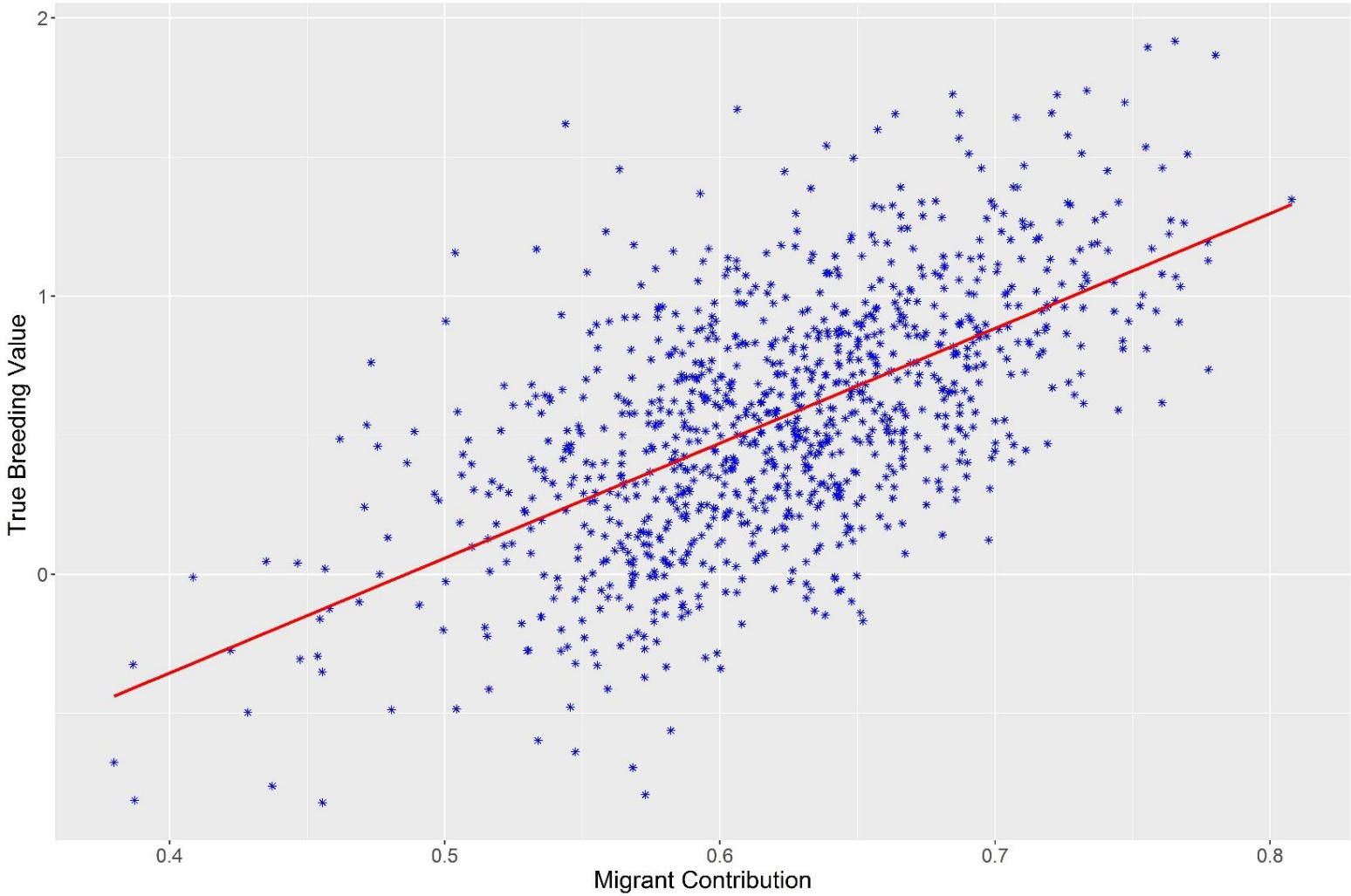
| **Scenario** | ***REF*** | | ***TS*** | | ***OCS-I*** | | ***OCS-II*** | | ***OCS-III*** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** |
| $G_0$ | 0.197 | 0.000 | 0.197 | 0.000 | 0.197 | 0.000 | 0.197 | 0.000 | 0.197 | 0.000 |
| $G_1$ | 0.197 | 0.000 | 0.112 | 0.005 | 0.120 | 0.002 | 0.124 | 0.002 | 0.121 | 0.003 |
| $G_2$ | 0.150 | 0.008 | 0.085 | 0.004 | 0.092 | 0.010 | 0.105 | 0.011 | 0.110 | 0.009 |
| $G_3$ | 0.124 | 0.005 | 0.075 | 0.003 | 0.082 | 0.012 | 0.094 | 0.014 | 0.095 | 0.007 |
| $G_4$ | 0.111 | 0.004 | 0.068 | 0.003 | 0.072 | 0.009 | 0.079 | 0.009 | 0.082 | 0.005 |
| $G_5$ | 0.102 | 0.006 | 0.062 | 0.003 | 0.074 | 0.011 | 0.069 | 0.005 | 0.078 | 0.010 |
| $G_6$ | 0.093 | 0.002 | 0.056 | 0.004 | 0.064 | 0.005 | 0.064 | 0.003 | 0.079 | 0.009 |
| $G_7$ | 0.093 | 0.002 | 0.055 | 0.002 | 0.059 | 0.005 | 0.062 | 0.002 | 0.075 | 0.004 |
| $G_8$ | 0.089 | 0.005 | 0.052 | 0.003 | 0.055 | 0.003 | 0.060 | 0.004 | 0.075 | 0.004 |
| $G_9$ | 0.091 | 0.003 | 0.049 | 0.002 | 0.052 | 0.004 | 0.059 | 0.003 | 0.071 | 0.005 |
| $G_{10}$ | 0.091 | 0.005 | 0.044 | 0.002 | 0.049 | 0.002 | 0.056 | 0.001 | 0.065 | 0.003 |

**Additional file 1 Table 4.S8** Descriptive statistics of the average genic variance ($\sigma^2_A$) achieved in each generation for each scenario

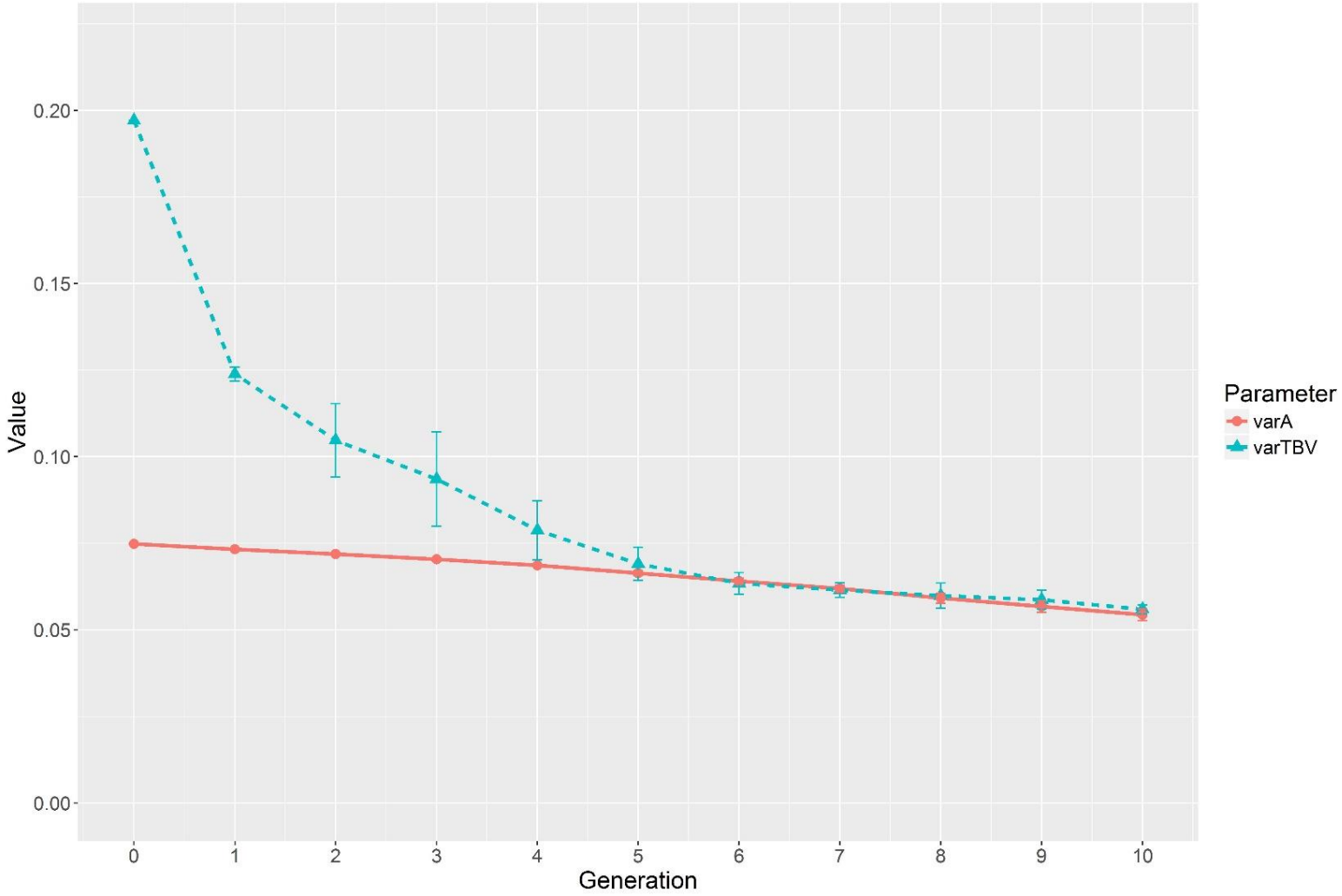
| **Scenario** | ***REF*** | | ***TS*** | | ***OCS-I*** | | ***OCS-II*** | | ***OCS-III*** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** |
| $G_0$ | 0.075 | 0.000 | 0.075 | 0.000 | 0.075 | 0.000 | 0.075 | 0.000 | 0.075 | 0.000 |
| $G_1$ | 0.075 | <0.001 | 0.073 | <0.001 | 0.073 | <0.001 | 0.073 | <0.001 | 0.073 | <0.001 |
| $G_2$ | 0.075 | <0.001 | 0.072 | <0.001 | 0.072 | <0.001 | 0.072 | <0.001 | 0.073 | <0.001 |
| $G_3$ | 0.075 | <0.001 | 0.070 | <0.001 | 0.070 | <0.001 | 0.070 | <0.001 | 0.071 | <0.001 |
| $G_4$ | 0.075 | <0.001 | 0.067 | <0.001 | 0.068 | 0.001 | 0.069 | <0.001 | 0.070 | <0.001 |
| $G_5$ | 0.075 | <0.001 | 0.064 | 0.001 | 0.065 | 0.001 | 0.066 | 0.001 | 0.069 | 0.001 |
| $G_6$ | 0.075 | <0.001 | 0.061 | 0.001 | 0.063 | 0.002 | 0.064 | 0.001 | 0.068 | 0.001 |
| $G_7$ | 0.075 | <0.001 | 0.058 | 0.001 | 0.060 | 0.002 | 0.062 | 0.001 | 0.067 | 0.001 |
| $G_8$ | 0.075 | <0.001 | 0.055 | 0.001 | 0.057 | 0.001 | 0.059 | 0.002 | 0.066 | 0.002 |
| $G_9$ | 0.075 | <0.001 | 0.052 | 0.001 | 0.054 | 0.002 | 0.057 | 0.002 | 0.065 | 0.002 |
| $G_{10}$ | 0.075 | <0.001 | 0.049 | 0.001 | 0.052 | 0.002 | 0.054 | 0.002 | 0.063 | 0.002 |




**Additional file 2 Figure 4.S1** Plot of the first two principal components for the dataset of the original population. The analysis was based on 889 individuals and 23,448 SNPs. Different colors and shapes represent individuals from different breeds




**Additional file 3 Figure 4.S2** Relationship between migrant contribution and true breeding values of Angler cattle in the simulated base population $G_0$. The correlation between migrant contribution and true breeding values is 0.599 and the regression coefficient is 4.131.




**Additional file 3 Figure 4.S3** Average genic variance ($\sigma^2_A$) and variance of true breeding values ($\sigma^2_{TBV}$) achieved in each generation of scenario OCS-II.

# CHAPTER 5

# Recovering the native genetic background of local breeds with historical introgression using advanced optimum contribution selection

Yu Wang[1*], Jörn Bennewitz[1], Robin Wellmann[1]

[1]Institute of Animal Science, University of Hohenheim, 70593 Stuttgart, Germany

[*]Corresponding author

Submitted for publication

# Abstract

Upgrading local livestock breeds with mainstream breeds has often proceeded over several decades and could eventually replace the native genetic background of these breeds, which would lead to their genetic extinction. De-introgression efforts aim at recovering their native genetic background while maintaining their contribution to the genetic diversity across breeds.

This study evaluated the long-term performance of different advanced optimum contribution selection strategies. Several scenarios with 25 generations of management and a population size of 400 were simulated based on the genotypes of German Angler cattle. The scenarios were compared at the time when the mean kinship surpassed a threshold value. The best scenario was able to increase the native contribution from 0.317 to 0.706 before a segment-based kinship level of 0.10 was reached. This scenario maximized the native contribution, constrained the increase in kinship, and the increase in kinship at native alleles. Moreover, it constrained the mean kinship in a multi-breed core set to the current level.

Recovering the native genetic background of an endangered breed removes alleles of mainstream breeds from its gene pool. Consequently, the maximum across-breed genetic diversity would be obtained by higher contributions of mainstream breeds to the multi-breed core set. This is at the expense of the endangered breed if it is not able to maintain a high genetic diversity. Consequently, the mean kinship in the core set needs to be restricted to avoid that the advantageous effect of making the breed more dissimilar to other breeds becomes overcompensated by the loss of its genetic diversity.

**Keywords**: de-introgression, conservation, optimum contribution selection, genetic diversity, runs of homozygosity

# 5.1 Background

For local livestock breeds, introgression with high yielding breeds was frequently implemented to improve performance (Hartwig *et al.* 2014, 2015). However, the gene-flow from high-yielding breeds to low-yielding breeds has an impact on the genetic diversity across breeds, and may also cause local breeds to lose their specific characteristics and adaptive traits, such as disease resistance and adaptation to a specific climate or harsh conditions (Taberlet *et al.* 2008). It often proceeded over several decades and could eventually replace the native genetic background of these breeds, which would lead to their genetic extinction. It is important to recover the native genetic background of local breeds to avoid their genetic extinction (Todesco *et al.* 2016), although the foreign genetic material is hard to remove after several generations of introgression (Amador *et al.* 2011).

A conservation program for a local breed with an introgression history should not only recover its native genetic background, but it should also maintain genetic diversity across breeds, which is affected by both within-breed diversity and between-breed diversity (Meuwissen 2009). It is defined with respect to a core set, which is a hypothetical set of individuals from several breeds. The percentage explained by each breed in the core set is determined such that the neutral gene diversity in the core set is maximized (Eding *et al.* 2002). Recovering the native genetic background may affect the genetic diversity in the core set because the breed becomes more inbred.

The basic approach for eliminating introgressed genetic material (de-introgression) is to detect the animals with a high proportion of the genome originating from native ancestors and to promote using them for breeding to produce the next generation (Toro *et al.* 2014). Amador *et al.* (2011) found that, based on pedigree information, the best strategy was to minimize the kinship between breeding candidates and the exogenous individuals that entered the population. Most of the recovery was achieved in the first generation of management when pedigree information was used. Genome-based strategies have better performance compared to the pedigree-based strategies since they detect the exogenous genetic material more efficiently (Amador *et al.* 2013).

Several factors may have an impact on the success in de-introgression: the genetic differentiation between the local breed and the populations it has mixed with, the percentage

of undesired background introgressed in the population, and the length of the admixture period. Based on the genome-wide information, the native genetic background could be fully recovered when the contribution of the foreign population is limited to 30-40% and the number of generations of admixture is not too high (1-3 generations) (Amador *et al.* 2013), but this is rarely the case. Although this strategy succeeded in removing foreign genetic material, reducing the genetic contributions from other breeds caused an increased inbreeding level, and thus reduced the within-breed diversity.

Traditional optimum contribution selection (traditional OCS) aims at maximizing genetic gain while controlling the rate of inbreeding by optimizing the genetic contribution of each selection candidate to the next generation (Meuwissen 1997; Woolliams *et al.* 2015). In our previous study (Wang *et al.* 2017a), we proposed the advanced OCS strategies that had the primary objective of maximizing genetic gain, but the loss of genetic originality was avoided by restricting the increase in kinship at native alleles and the native contribution (NC) in the OCS process. It was shown that genetic progress was substantially lower in scenarios that additionally aimed at recovering the original genetic background. Hence, constraining native contribution at their current values was recommended to avoid a further reduction of genetic originality. For some breeds, breeding values may be less important or not available. For these breeds, the primary objective could be recovering the native genetic background. This breeding objective is evaluated in the present study.

In the previous studies on de-introgression (Amador *et al.* 2011, 2013), the level of genetic diversity was only monitored and evaluated within the breed of interest, but the impact of de-introgression on the total genetic diversity of the core set was not studied. Therefore, the objective of the present study was to evaluate the long-term performance of different genomic advanced OCS strategies that aim at recovering the native genetic background of a breed with historical introgression. This was done either by maximizing the native contribution or by minimizing the kinship in the multi-breed core set. The increase in inbreeding and the increase in inbreeding at native alleles were restricted in most scenarios. The performances were evaluated based on results from several subsequent simulated generations.

# 5.2 Material and methods

## 5.2.1 Simulation

The target breed for conservation in this study is the German Angler Rotvieh cattle (shortened as Angler). It is a dual purpose breed with an emphasis on milk production and mainly located in Northern part of Germany (Bennewitz *et al.* 2008). In addition, a multi-breed dataset was considered consisting of 400 Angler, 200 Holstein-Friesian, 200 Red Holstein, 200 Fleckvieh and 21 Norwegian Red cattle to evaluate the genetic diversity in a multi-breed core set in each generation. The management of the population was only performed for Angler cattle and the animals from the other breeds remain the same in each generation. The genotypes of the other breeds were also used for identifying native alleles of Angler cattle. The genotype information we used in this study was part of the dataset of our previous study (Wang *et al.* 2017b).

The simulation process consisted of two parts. First, a base population ($G_0$) consisting of 400 Angler animals (200 male and 200 female animals) was generated from 123 Angler sires and 132 dams. Afterwards, 25 subsequent generations were simulated and managed according to different scenarios. All animals have genomic information, whereby 23,448 SNPs were used for the analysis. Haplotypes were phased for all breeds jointly in a larger data set and missing genotypes were imputed using BEAGLE software (Browning and Browning 2007). The detail of the simulation process can be found in Wang *et al.* (2017b). Selection started in generation $G_0$. The number of offspring of each parent was determined by calculating its optimum genetic contribution $c_i$ to the next generation. Matings were randomly assigned. Offspring received the haplotypes from the parents via Mendelian inheritance. Crossovers occurred with 1% probability per cM and were uniformly distributed along the chromosomes (Weng *et al.* 2014). The population size of Angler cattle remained 400 every generation (200 male and 200 female animals).

## 5.2.2 Genetic parameters

Four genetic parameters were monitored: the mean native contribution (NC), the mean kinship across breeds in the core set ($f_{CORE}$), the mean kinship of Angler cattle ($f_{SEG}$) and the mean kinship at native alleles of Angler cattle ($f_{SEG|N}$).

The allele origin at each position in each Angler haplotype was determined in generation $G_0$ with R package *optiSel* v 2.0 (Wellmann 2018). A haplotype was classified to be native at a particular marker position if the frequency of the segment containing the marker is sufficiently low in all other breeds. The threshold for the frequency was set as 0.01 in this study. A segment was required to consist of at least 20 consecutive markers with a minimum length of 1.5Mb. The origins of the alleles were traced through the following generations and the native contribution of each Angler individual was calculated as the proportion of the genome that was classified to be native in Angler.

In all the optimization scenarios, we considered a hypothetical multi-breed population for monitoring the level of genetic diversity across breeds (Wellmann *et al.* 2014). Except for the Angler cattle from the respective generation, animals from Holstein-Friesian, Red Holstein, Fleckvieh and Norwegian Red were also included in the core set. In each scenario, the contribution of each breed to the core set was determined that minimized the segment-based kinship in the core set, which is equivalent to maximizing its genetic diversity. Since we aimed to maximize the diversity of alleles across the breeds, we used the diversity measures of Eding *et al.* (2002), which assigned equal weight to both within and between breed diversity. The average kinship across breeds ($f_{CORE}$) is calculated as $\mathbf{bc'f_{CORE}bc}$, where $\mathbf{bc}$ is the vector of breed contributions to the multi-breed core set and $\mathbf{f_{CORE}}$ is the 5x5 matrix of average kinships within and across breeds.

Two breed specific kinship parameters were considered during OCS, which are the segment-based kinship $f_{SEG}$ and the segment-based kinship at native alleles $f_{SEG|N}$ in Angler cattle. Kinship $f_{SEG}$ between individual i and j (element of the matrix $\mathbf{f_{SEG}}$) is the probability that two randomly chosen alleles from two individuals belong to a shared segment (de Cara *et al.* 2013). The average kinship of Angler in the subsequent generation is predicted as $\mathbf{c'f_{SEG}c + l_{SEG}(c)}$, where $\mathbf{c}$ is the vector of the genetic contributions of the selection candidates from the Angler breed and $\mathbf{l_{SEG}(c)}$ is a linear correction term that accounts for genetic drift (Wellmann 2018). For recovering the genetic background of a breed with historical introgression, Wellmann *et al.* (2012) proposed to take the kinship at native alleles $f_{SEG|N}$ into consideration. When computed from genotypes, this is the conditional probability that two randomly chosen alleles from two Angler individuals belong to a shared segment, given that they are native. The corresponding mathematical definitions can be found in Wang *et al.* (2017b).

### 5.2.3 Basic assumptions for OCS scenarios

Several assumptions were made for all scenarios: The genetic contribution of each selection candidate is non-negative ($c_i \geq 0$); the sum of genetic contribution from each sex group equals to 0.5: $\mathbf{c}'\mathbf{s} = 0.5$ and $\mathbf{c}'\mathbf{d} = 0$ where $\mathbf{s}$ and $\mathbf{d}$ are vectors indicating candidates' sex. The optimization process was done for both male and female selection candidates. For male selection candidates, the number of offspring was not restricted, so the maximum genetic contribution of each male selection candidate is 0.5 ($c_{\mathrm{sire}_i} \leq 0.5$). Cows can only produce a limited number of offspring per generation and the maximum genetic contribution for female selection candidates was set to 0.01 ($c_{\mathrm{dam}_i} \leq 0.01$), which resulted in at most eight offspring per cow.

### 5.2.4 Objective functions and scenarios

We considered nine scenarios aiming at recovering the native genetic background of Angler cattle and at increasing the between-breed genetic diversity. One additional reference scenario (REF) was used for comparison. Two objective functions were considered: maximizing the native contribution in each following generation, and minimizing the average kinship in the multi-breed core set in each following generation, whereby only the contributions of Angler cattle were optimized. The settings of each scenario are shown in Table 5.1. Five replicates per scenario were simulated and the results presented are averages across replicates.

***Reference scenario (REF)***

All animals were used in this scenario as parents. No optimization and selection was done, so every individual had two offspring.

***Maximum native contribution (MNC)***

The genetic contribution of each selection candidate was optimized by maximizing the native contribution at each following generation ($\mathbf{c}'\mathbf{NC}$). Except for the basic assumptions mentioned above, no additional constraints were considered.

***Scenarios aiming at maximizing native genetic contribution ($maxNative.*$)***

In these scenarios, the mean native contribution at each following generation ($\mathbf{c'NC}$) was maximized, while the average kinship $f_{SEG}$ was constrained in order to restrict the rate of inbreeding. Four scenarios with a different combination of additional constraints were constructed, which are named maxNative.A, maxNative.B, maxNative.C and maxNative.D. The additional constraints, which are listed in Table 5.1, restricted $f_{SEG|N}$ and $f_{CORE}$.

***Scenarios aiming at minimizing kinship across breeds ($minf_{CORE}.*$)***

In these scenarios, the average kinship in the multi-breed core set at each following generation was minimized, while the average kinship $f_{SEG}$ in Angler was constrained. Additional constraints restricted NC and $f_{SEG|N}$ in some scenarios. This resulted in four scenarios, which are named $minf_{core}.A$, $minf_{core}.B$, $minf_{core}.C$ and $minf_{core}.D$. The constraint settings can be found in Table 5.1.

**Table 5.1** Scenario description

| **Scenario name** | **Objective function** | **Constraints[1]** |
|---|---|---|
| REF | Reference scenario (no selection & optimization) | No constraint |
| MNC | maximize native contribution ($\mathbf{c'NC}$) | Common constraints[2] |
| $maxNative.A$ | maximize native contribution ($\mathbf{c'NC}$) | $ub.f_{SEG}$ |
| $maxNative.B$ | | $ub.f_{SEG}, ub.f_{CORE}$ |
| $maxNative.C$ | | $ub.f_{SEG}, ub.f_{SEG|N}$ |
| $maxNative.D$ | | $ub.f_{SEG}, ub.f_{SEG|N}, ub.f_{CORE}$ |
| $minf_{CORE}.A$ | minimize the kinship across breeds ($\mathbf{c'f_{CORE}c}$) | $ub.f_{SEG}$ |
| $minf_{CORE}.B$ | | $ub.f_{SEG}, lb.NC$ |
| $minf_{CORE}.C$ | | $ub.f_{SEG,} ub.f_{SEG|N}$ |
| $minf_{CORE}.D$ | | $ub.f_{SEG,} ub.f_{SEG|N}, lb.NC$ |

[1] Upper (ub.*) or lower (lb.*) limit of the parameter in the corresponding scenario. NC: native contribution; $f_{CORE}$: average kinship across the breeds; $f_{SEG}$: average kinship of Angler cattle; , $f_{SEG|N}$: average kinship at native alleles of Angler cattle.

[2]Common constraints were applied for all the optimized scenarios: breed contribution was optimized for maximize genetic diversity of the multi-breed population; no offspring number limitation for male selection candidates; maximum genetic contribution for female selection candidates is 0.01.

### 5.2.5 Constraint values

To restrict the rate of inbreeding, we assumed that an effective population size of at least $N_e$=75 needs to be maintained. According to $N_e = 1/(2\Delta F)$, the desired rate of inbreeding was 0.667% per generation (Falconer and Mackay 1996). Thus, the upper bound for the mean kinship in Angler ($f_{SEG}$) in the next generation was set as

$$\text{ub.}\, f_{SEG_{t+1}} = \overline{f_{SEG_t}} + \left(1 - \overline{f_{SEG_t}}\right) \Delta F$$

where $\overline{f_{SEG_t}}$ is the average kinship of Angler cattle in the current generation t. Similarly, the upper bound for the mean native kinship in Angler ($f_{SEG|N}$) in the next generation was

$$\text{ub.}\, f_{SEG|N_{t+1}} = \overline{f_{SEG|N_t}} + \left(1 - \overline{f_{SEG|N_t}}\right) \Delta F$$

where $\overline{f_{SEG|N_t}}$ is the average kinship at native alleles of Angler cattle in the current generation t. For scenarios constraining $f_{CORE}$, we assumed that in each generation, the average kinship in the core set did not exceed the average kinship in generation $G_0$. For scenarios constraining the native contribution, the lower bound for the mean NC in the next generation was defined as the 65% quantile of the NCs in the current generation. That is,

$$\text{lb.}\, NC_{t+1} = \text{quantile}(NC_t, 0.65)$$

For all optimization scenarios, solver "cccp" (Pfaff 2014), which was called from R package *optiSel* v 2.0 (Wellmann 2018), was used to solve the optimization problems. Five replicates per scenario were simulated and the results presented are averages across these replicates.

# 5.3 Results

## 5.3.1 Values of each parameter obtained in all scenarios

The mean and standard deviation of each parameter at the beginning of selection ($G_0$) and the end of selection ($G_{25}$) are shown in Table 5.2 for all scenarios. The values were estimated based on five replicates. The values of the parameters for all generations can be found in Table 5.S1-5.S9 [Additional File 5.S1: SupplementTables.xlsx].

The average NC at the beginning of selection is 0.3171. All scenarios aiming at maximizing NC had good performance in achieving high native contribution after 25 generations of management, which is shown in Figure 5.1. Scenario MNC achieved higher NC compared to the other scenarios until generation $G_{15}$. Afterwards, the increase of native contribution became slower and finally reached a plateau (0.6975). Simultaneously due to the absence of controlling the kinship among the selection candidates in MNC, $f_{SEG}$ increased to 0.9422 and $f_{SEG|N}$ increased to 0.9660 in generation $G_{25}$. Among all scenarios aiming at maximizing NC that restricted $f_{SEG}$, scenario maxNative.A achieved the highest NC level with a mean value of 0.7870 in generation $G_{25}$, but the kinship across breeds increased from 0.0285 in generation $G_0$ to 0.0308 after 25 generations' management. The inclusion of $f_{CORE}$ and/or $f_{SEG|N}$ as additional constraint automatically reduced the level of $f_{SEG}$ in scenario maxNative.B (0.1122), maxNative.C (0.1254) and maxNative.D (0.1025) below the constraint setting.

The trend of average kinship across breeds of scenario REF, MNC and scenarios aiming at minimizing $f_{CORE}$ can be seen in Figure 5.2. REF had a stable $f_{CORE}$ level from $G_0$ (0.0285) to $G_{25}$ (0.0288). MNC resulted in the highest $f_{CORE}$ among all scenarios at the end of selection (0.0322) because both $f_{SEG}$ and $f_{CORE}$ were not restricted in this scenario. Scenarios $minf_{CORE}$.A (0.0237) and $minf_{CORE}$.C (0.0239) had good performance in reducing the kinship across breeds. However, both scenarios had little function in increasing the native genetic contribution in Angler, which only slightly increased to 0.366. When the native contribution was constrained, the $f_{CORE}$ level increased to 0.0251 in $minf_{CORE}$.B and 0.0253 in $minf_{CORE}$.D. All scenarios aiming at minimizing $f_{CORE}$ maintained a low kinship $f_{SEG}$ (0.0477 – 0.0586) and a low kinship at native alleles $f_{SEG|N}$ (0.0728 – 0.0904).

**Table 5.2** Basic statistics of each parameter achieved in the base generation $G_0$ and in $G_{25}$ for each selection scenario

| | Parameter[1,2] | | | |
|---|---|---|---|---|
| | NC | $f_{CORE}$ | $f_{SEG}$ | $f_{SEG\|N}$ |
| Beginning of selection ($G_0$) | | | | |
| | 0.3171 ± 0.0000 | 0.0285 ± 0.0000 | 0.0546 ± 0.0000 | 0.0763 ± 0.0000 |
| End of selection ($G_{25}$)[2] | | | | |
| REF | 0.3189 ± 0.0048 | 0.0288 ± 0.0001 | 0.0625 ± 0.0010 | 0.0882 ± 0.0042 |
| MNC | **0.6975 ± 0.0054** | 0.0322 ± 0.0000 | 0.9422 ± 0.0104 | 0.9660 ± 0.0071 |
| maxNative. A | **0.7870 ± 0.0092** | 0.0308 ± 0.0001 | *0.1904 ± 0.0025* | 0.2932 ± 0.0056 |
| maxNative. B | **0.7337 ± 0.0096** | *0.0285* ± 0.0000 | *0.1122 ± 0.0010* | 0.1728 ± 0.0022 |
| maxNative. C | **0.7415 ± 0.0040** | 0.0301 ± 0.0001 | *0.1254 ± 0.0016* | *0.1714 ± 0.0017* |
| maxNative. D | **0.7243 ± 0.0060** | *0.0285* ± 0.0000 | *0.1025 ± 0.0011* | *0.1584 ± 0.0018* |
| $minf_{CORE}$. A | 0.3665 ± 0.0065 | **0.0237 ± 0.0000** | *0.0478 ± 0.0003* | 0.0728 ± 0.0007 |
| $minf_{CORE}$. B | *0.5617 ± 0.0226* | **0.0251 ± 0.0004** | *0.0586 ± 0.0037* | 0.0904 ± 0.0064 |
| $minf_{CORE}$. C | 0.3661 ± 0.0021 | **0.0239 ± 0.0001** | *0.0477 ± 0.0006* | *0.0730 ± 0.0010* |
| $minf_{CORE}$. D | *0.5528 ± 0.0147* | **0.0253 ± 0.0003** | *0.0581 ± 0.0022* | *0.0895 ± 0.0037* |

[1]Parameters estimated in each generation of each scenario. NC: native contribution; $f_{CORE}$: average kinship across the breeds; $f_{SEG}$: average kinship of Angler cattle; , $f_{SEG|N}$: average kinship at native alleles of Angler cattle.

[2]Results are based on five replicates. The objective function is marked as bold in the scenario. The parameter used as a constraint is marked as italic in the scenario.

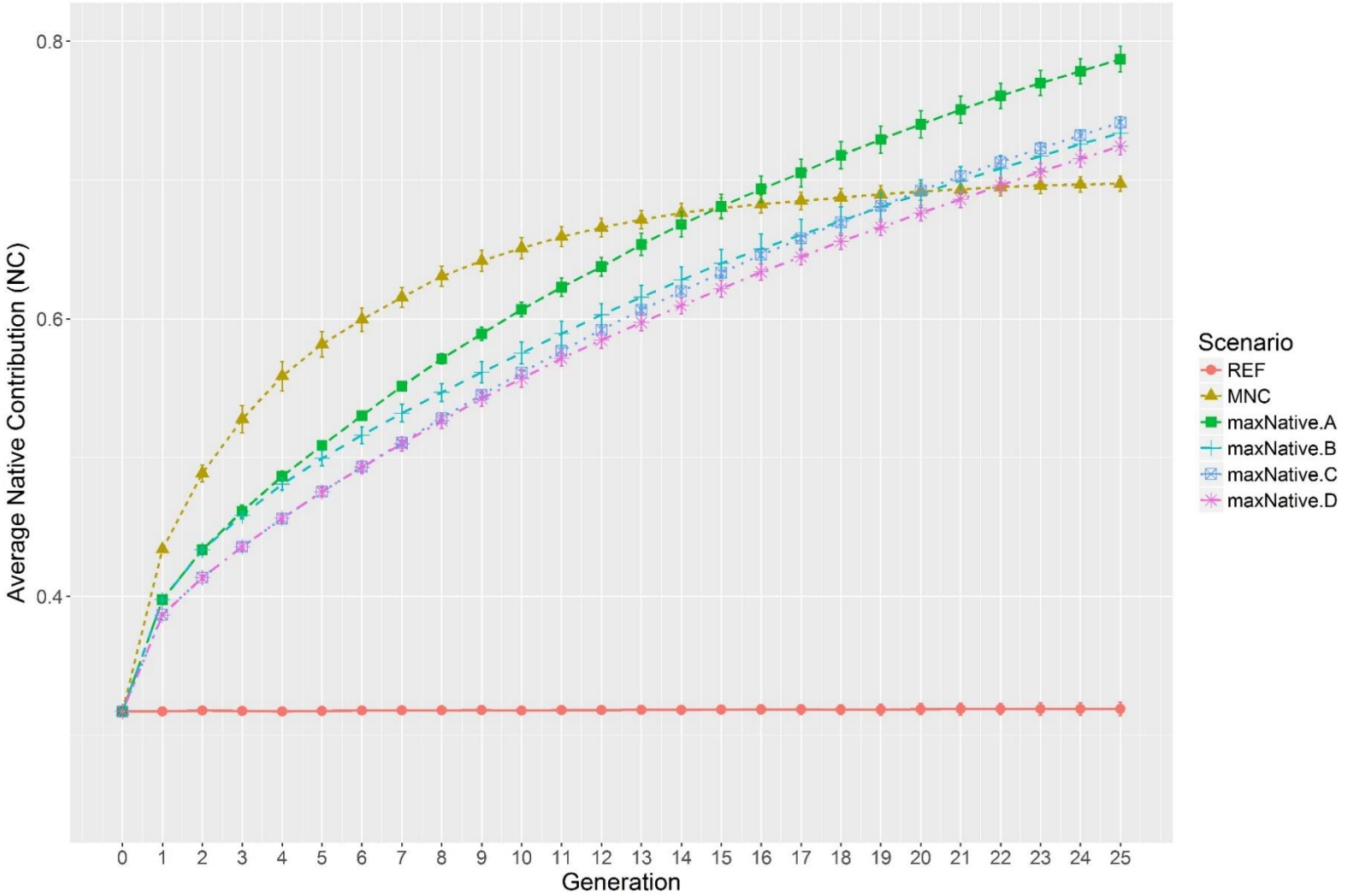



**Figure 5.1** Average native contribution achieved in each generation of each selection scenario

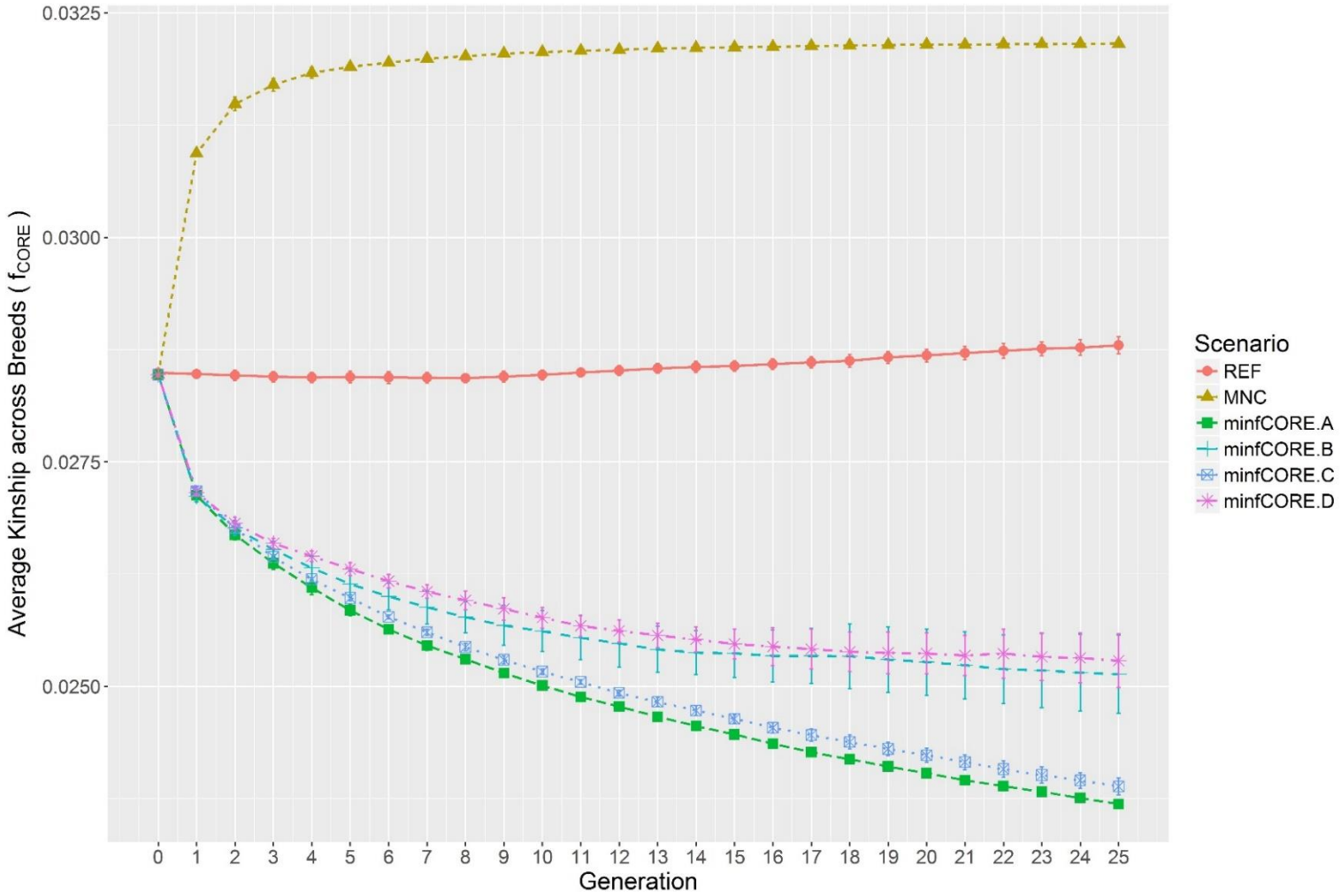



**Figure 5.2** Average kinship across breeds achieved in each generation of each selection scenario

### 5.3.2 The relationship between NC and $f_{CORE}$

The changes of the two targeted parameters NC and $f_{CORE}$ during selection are shown in the trajectory plot of Figure 5.3 for all generations and all scenarios. Both parameters were rather stable for REF as no trend can be observed. Reducing $f_{CORE}$ increased NC for all scenarios aiming at minimizing $f_{CORE}$. In contrast, increasing NC also increased $f_{CORE}$ in scenario MNC. After the first generation, the same could be observed for all other scenarios aiming at maximizing NC. However, increasing NC in scenarios maxNative.A and maxNative.B slightly decreased $f_{CORE}$ in the first generation.

We also compared the parameter values achieved in the scenarios at the generation in which the kinship of Angler cattle approached the threshold 0.10 (Table 5.3). Scenario MNC is not shown since $f_{SEG}$ increased to 0.1707 at $G_1$. All scenarios aiming at minimizing $f_{CORE}$ achieved a low $f_{SEG}$ ranging from 0.0478 to 0.0586, so the threshold was not reached. Scenario maxNative.D achieved the highest NC (0.7059) when $f_{SEG}$ approached 0.10, whilst the $f_{CORE}$ was maintained at the constrained level. In addition, for scenario maxNative.D, it took 23 generations for $f_{SEG}$ to increase to 0.10. This is partly because the segment-based kinship ignores very old inbreeding.

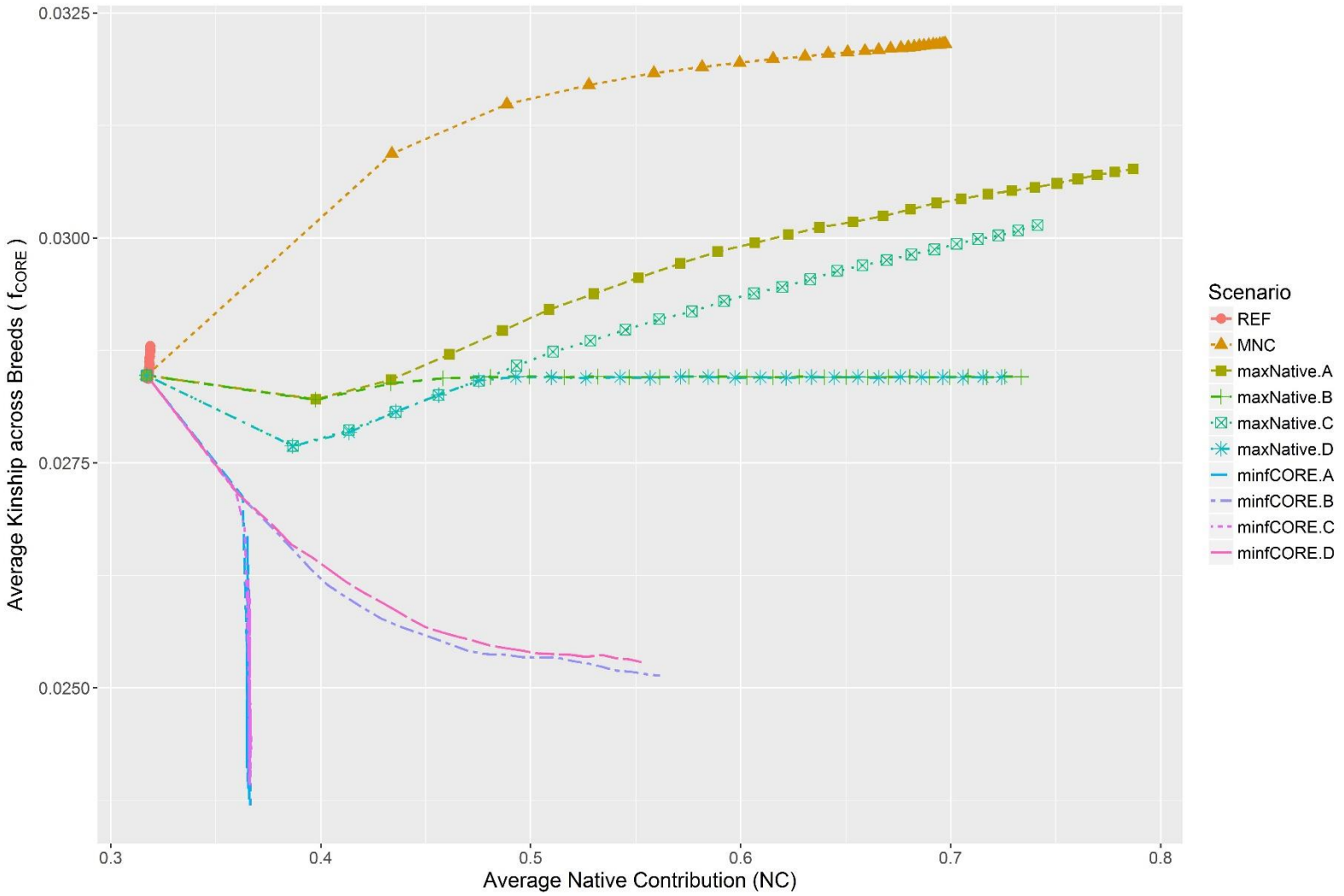



**Figure 5.3** The relationship between the average native contribution and kinship across breeds based on the results of all scenarios through the 25 generations' selection process.

**Table 5.3** Parameter values when the average kinship of Angler approached 0.10, otherwise at the last generation

| Scenario | Generation | NC[1] | $f_{CORE}$ | $f_{SEG}$ | $f_{SEG|N}$ |
|---|---|---|---|---|---|
| REF | $G_{25}$ | 0.3189 | 0.0288 | 0.0625 | 0.0882 |
| maxNative. A | $G_{7}$ | 0.5514 | 0.0296 | 0.0960 | 0.1685 |
| maxNative. B | $G_{17}$ | 0.6605 | 0.0285 | 0.0993 | 0.1618 |
| maxNative. C | $G_{16}$ | 0.6460 | 0.0296 | 0.0972 | 0.1411 |
| maxNative. D | $G_{23}$ | 0.7059 | 0.0284 | 0.0992 | 0.1551 |
| $minf_{CORE}$. A | $G_{25}$ | 0.3665 | 0.0237 | 0.0478 | 0.0728 |
| $minf_{CORE}$. B | $G_{25}$ | 0.5617 | 0.0251 | 0.0586 | 0.0904 |
| $minf_{CORE}$. C | $G_{25}$ | 0.3661 | 0.0239 | 0.0477 | 0.0730 |
| $minf_{CORE}$. D | $G_{25}$ | 0.5528 | 0.0253 | 0.0581 | 0.0895 |

[1]Parameters estimated in each generation of each scenario. NC: native contribution; $f_{CORE}$: average kinship across the breeds; $f_{SEG}$: average kinship of Angler cattle; , $f_{SEG|N}$: average kinship at native alleles of Angler cattle.

### 5.3.3 Breed contribution

In each generation of each scenario, a hypothetical multi-breed core set was considered. The contribution of each breed to the core set was computed such that the average kinship in the core set was minimized. The average breed contributions to the core set at the beginning of selection ($G_0$) and end of selection ($G_{25}$) are shown in Table 5.4. In addition, the changing trend of breed proportion of each generation is shown in Figure 5.S1 [Supplemental File 5.S2: Figure 5.S1.jpg]. The average breed contribution of Angler was 0.2773 in $G_{25}$ in REF. All scenarios aiming at maximizing native genetic contribution achieved a lower value compared to REF in $G_{25}$. The lowest value was close to zero, which was obtained in scenario MNC (0.0190). In contrast, in scenarios aiming at minimizing the kinship in the core set, Angler cattle reached a relatively high final contribution (0.3262 – 0.3863). In all scenarios, the breed contribution of Fleckvieh and Norwegian Red cattle are higher compared to the Holstein breeds.

**Table 5.4** Optimized breed contribution to the hypothetical multiple-breed core set to achieve maximum neutral gene diversity in base generation $G_0$ and $G_{25}$ for each selection scenario

| | Breed Contribution | | | | |
|---|---|---|---|---|---|
| | Angler | Fleckvieh | Holstein-Friesian | Red Holstein | Norwegian Red |
| Beginning of selection ($G_0$) | | | | | |
| | 0.3918 (REF) | 0.1959 (REF) | 0.1959 (REF) | 0.1959 (REF) | 0.0206 (REF) |
| | 0.3301 (Other) | 0.3642 (Other) | 0.0939 (Other) | 0.0048 (Other) | 0.2071 (Other) |
| End of selection ($G_{25}$) | | | | | |
| REF | 0.2773 | 0.3706 | 0.1090 | 0.0204 | 0.2227 |
| MNC | 0.0190 | 0.4234 | 0.1927 | 0.0715 | 0.2935 |
| maxNative. A | 0.0952 | 0.4023 | 0.1725 | 0.0603 | 0.2697 |
| maxNative. B | 0.1817 | 0.3700 | 0.1544 | 0.0498 | 0.2441 |
| maxNative. C | 0.1387 | 0.3929 | 0.1613 | 0.0510 | 0.2561 |
| maxNative. D | 0.1918 | 0.3698 | 0.1520 | 0.0460 | 0.2404 |
| $\text{minf}_{\text{CORE}}$. A | 0.3863 | 0.3029 | 0.1045 | 0.0200 | 0.1862 |
| $\text{minf}_{\text{CORE}}$. B | 0.3262 | 0.3233 | 0.1211 | 0.0263 | 0.2031 |
| $\text{minf}_{\text{CORE}}$. C | 0.3858 | 0.3051 | 0.1047 | 0.0184 | 0.1861 |
| $\text{minf}_{\text{CORE}}$. D | 0.3262 | 0.3250 | 0.1184 | 0.0277 | 0.2027 |

# 5.4 Discussion

In our previous study (Wang *et al.* 2017b), we proposed the advanced optimum contribution selection strategies to balance the needs for production and conservation to make sure that the local breed benefits from the improved economic performance. In this study, the main purpose was to increase the value of a breed for conservation by removing exogenous genetic material, maintaining within-breed genetic diversity, and increasing the genetic diversity among breeds. Production performance was not considered. The results showed that with a set of appropriate constraint settings, it is possible to achieve a large amount of recovery at a relatively low rate of inbreeding and to increase genetic diversity across breeds. Multiple generations of management are required via using the advanced OCS strategies.

## 5.4.1 Maximizing the native contribution

One approach to increase the value of a breed for conservation could be maximizing the native contribution. The greediest approach is to maximize the native contribution without constraining kinships (scenario MNC). This resulted in a very high rate of inbreeding as only the "purest" male selection candidate with the highest native contribution was used in each generation. Moreover, the average NC reached a plateau after about 15 generations. In comparison, NC continued to increase in all the other scenarios that aimed at maximizing the native contribution. Consequently, the greediest approach achieved the lowest recovery of NC after 25 generations. All other scenarios constrained the increase in mean kinship to achieve an effective population size of at least 75. Consequently, no plateau could be observed, but the increase in NC per generation was reduced in later generations.

Scenario maxNative.A, which only restricted the increase in mean kinship, achieved the highest recovery of the native background after 25 generations. It might be expected that maximizing the NC also reduces the mean kinship in the core set. This was, however, not the case. The low effective population size caused a reduction of genetic diversity in Angler, which counteracted the advantageous effect that Angler became more different from the other breeds. If more breeds would be included in the core set, the within-breed diversity would contribute less to the genetic diversity of the core set. Hence under this circumstance, maximizing NC is expected to be more desirable for increasing the across-breed diversity.

As the mean NC achieved at generation $G_{25}$ was greater in scenario maxNative.A than in maxNative.C, it follows that restricting the increase in $f_{SEG|N}$ at native alleles is a stronger constraint than restricting the increase in $f_{SEG}$ when both parameters are restricted to increase at the same speed, which is in accordance with our previous studies (Wang *et al.* 2017a, b). However, after about 15 generations of management, the increase in NC per generation became larger for scenario maxNative.C, and maintaining a high diversity of native alleles is a goal in itself.

Scenarios maxNative.A and maxNative.C both reduced the genetic diversity in the core set because the diversity within the breed decreased. Consequently, an additional constraint should be applied which causes the diversity in the core set not to decrease. Compare to the other optimized scenarios, maxNative.D achieved a slightly lower NC level. However, it resulted in a considerably larger effective population size, a considerably smaller decrease in the genetic diversity of native alleles, and conserved the genetic diversity in the core set. Hence, if the priority of a breeding program is to recover the native genetic background, then scenario maxNative.D is the recommended approach, but more breeds should be included in the core set.

### 5.4.2 Minimizing the kinship in the core set

An alternative to recover the native genetic background of a breed is to make the breed genetically more dissimilar to other breeds. We tried to accomplish this goal by minimizing the average kinship in the core set. Scenario $minf_{CORE}$.A shows that the mean kinship in the core set could be reduced from 0.0285 in the beginning to 0.0237 at generation $G_{25}$, whereby selection was only in Angler cattle and the contribution of the Angler cattle to the core set increased from 0.330 to 0.386. This means that Angler cattle become more valuable for conservation. This is, however, primarily achieved by increasing genetic diversity in Angler. The native contribution in Angler increased only had a small increase. All scenarios that minimized the kinship in the core set resulted in an effective population size that was considerably larger than 75, which was the envisaged value. Hence, minimizing the kinship in the core set indeed increases the value of a breed for conservation, but this is primarily achieved by increasing its genetic diversity, and only little by making the breed genetically more dissimilar to other breeds. This would be less pronounced for a breed whose contribution to the core set is smaller, and thus also in the case that more breeds are included in the core set.

### 5.4.3 Genetic progress in NC until a particular inbreeding level is reached

In a conservation program, the time needed to achieve a breeding goal may be less important than the genetic progress that can be achieved until the inbreeding level surpasses a given threshold value. Therefore we compared the scenarios at the generation in which the mean kinship approached the threshold value 0.10. This was the case in generation $G_7$ for scenario maxNative.A, but it was not even achieved in generation $G_{25}$ in the scenarios that minimized the mean kinship in the core set. It can be seen in Table 5.3 that the highest recovery of the genetic background could be achieved in scenario maxNative.D. This also shows that scenario maxNative.D is the most appropriate method for recovering the native background.

### 5.4.4 Breed contribution to the between-breed genetic diversity

In our previous study, we found that Angler cattle had a close relationship with Red Holstein and Holstein, a moderately close relationship with Norwegian Red, and a distant relationship with the Fleckvieh (Wang *et al.* 2017b). Thus, it would be expected that recovering the native genetic background in Angler removes Holstein alleles from Angler. Consequently, Holstein genes would become rarer in the core set, if this is not counteracted by an increase of the contributions of Holstein and Red Holstein to the core set. Figure 5.S1 shows that this would be indeed the case. In particular, as it showed in Figure 5.4, for scenario maxNative.D, which kept the kinship in the core set constant, shows how recovering the native genetic background of an endangered breed increases the contributions of the mainstream breeds to the core set as their genes are no longer present in the endangered breeds. This is done at the expense of the contribution of the endangered breed if it is not able to maintain a large effective population size.

The final contribution of Angler to the core set is lower in scenarios aiming at maximizing the native genetic contribution of Angler than in scenarios aiming at minimizing the kinship across breeds. Especially in the extreme case of scenario MNC, after 25 generations of management, Angler cattle contributed only 1.9% to the core set, indicating that whether including Angler or not has little impact on the between-breed diversity. In all cases, the reduced contributions of Angler cattle may have resulted from the reduction of within-breed diversity in the de-introgression process.

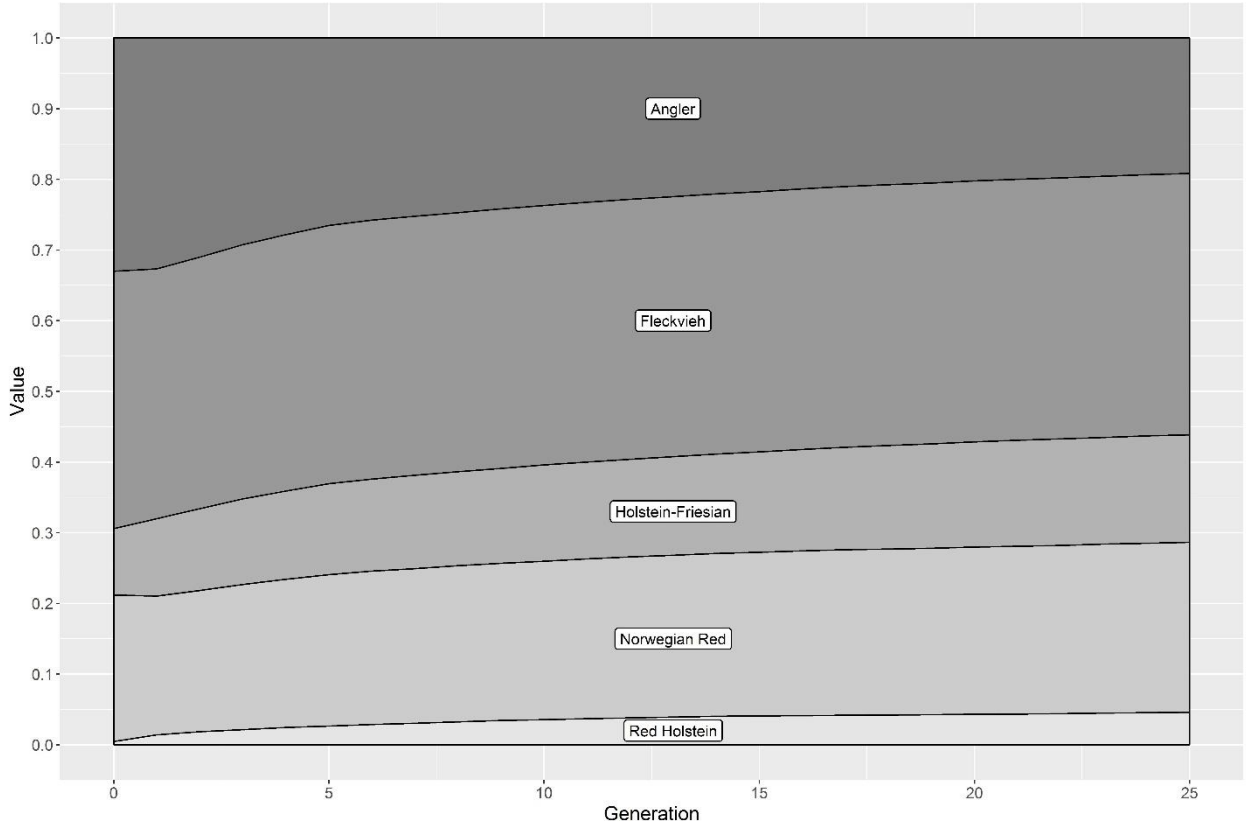



**Figure 5.4** Breed contribution to the hypothetical multi-breed population to maximize neutral gene diversity in scenario maxNative.D

## 5.4.5 Future improvement

It was shown that the native contribution can be increased considerably in a population even if a low rate of inbreeding is maintained. This required, however, many generations of selection. The time required would be shorter for a larger population because of a higher selection intensity. A higher selection intensity could also be achieved if embryo transfer is used routinely and if only genotyped embryos with high native contribution are carried out.

In this study, we aimed at maintaining or increasing the gene diversity of the core set. This is expected to maintain the ability of the species to adapt to new environments in the long term (Wellmann *et al.*, 2014). However, if only a few generations of selection are permissible to adopt a breed towards a new breeding goal, then the maintenance of adaptive diversity may be more important than the maintenance of gene diversity. This was not considered in our paper.

Instead of maximizing gene diversity, several studies have provided suggestions in putting different weights on within-breed diversity and between-breed diversity for conservation

decisions (d'Arnoldi *et al.* 1998; Eding and Meuwissen 2001; Piyasatian and Kinghorn 2003; Bennewitz and Meuwissen 2005). Toro *et al.* (2006) suggested an intermediate value for the weight on within-breed diversity in order to not prioritize neither small inbred lines nor large non-endangered breeds for conservation. Indeed, a lower weight could be put on within-breed diversity, since the conservation of within-breed diversity can be achieved by a different constraint.

# 5.5 Conclusion

The value of an endangered breed with historic introgression for conservation could be increased by recovering the native genetic background. In a conservation program, the time needed for recovering the native genetic background may be less important than the genetic progress that can be achieved until the inbreeding level surpasses a given threshold value. The highest recovery was achieved by scenario maxNative.D, which was able to increase the native contribution from 0.317 to 0.706 before a segment-based kinship level of 0.10 was reached. This scenario maximized the native contribution and constrained the kinship, the kinship at native alleles and the across-breed diversity. The native contribution recovered after 25 generations by this scenario was 0.724.

Increasing the native contribution in an endangered breed could change the contributions of all breeds to the core set, which is a hypothetical multi-breed population with maximum genetic diversity. In particular, the contributions of the mainstream breeds, whose genes become removed from the endangered breeds, are increasing. This takes place at the expense of the contributions of the endangered breed if it is not able to maintain a large effective population size. Consequently, the mean kinship in the core set needs to be constrained not to increase during the recovery of the native genetic background of the endangered breed. Otherwise, the positive effect of making the breed more dissimilar to other breeds would be overcompensated by the loss of genetic diversity within the breed. Increasing genetic diversity in the core set by advanced OCS in an endangered breed would primarily be achieved by increasing the genetic diversity within the breed and only little by making it more dissimilar to other breeds.

## Acknowledgments

The study was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG).

## Competing interests

The authors declare that there is no conflict of interest regarding the publication of this article

# References

Amador C., Toro M. Á., Fernández J., 2011 Removing exogenous information using pedigree data. Conserv. Genet. 12: 1565–1573.

Amador C., Fernández J., Meuwissen T. H. E., 2013 Advantages of using molecular coancestry in the removal of introgressed genetic material. Genet. Sel. Evol. 45: 13.

Bennewitz J., Meuwissen T. H. E., 2005 A novel method for the estimation of the relative importance of breeds in order to conserve the total genetic variance. Genet. Sel. Evol. 37: 315–337.

Bennewitz J., Simianer H., Meuwissen T. H. E., 2008 Investigations on merging breeds in genetic conservation schemes. J. Dairy Sci. 91: 2512–2519.

Browning S. R., Browning B. L., 2007 Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81: 1084–1097.

d'Arnoldi C. T., Foulley J.-L., Ollivier L., 1998 An overview of the Weitzman approach to diversity. Genet. Sel. Evol. 30: 149–161.

de Cara M. Á. R., Villanueva B., Toro M. Á., Fernández J., 2013 Using genomic tools to maintain diversity and fitness in conservation programmes. Mol. Ecol. 22: 6091–6099.

Eding H., Meuwissen T. H. E., 2001 Marker-based estimates of between and within population kinships for the conservation of genetic diversity. J. Anim. Breed. Genet. 118: 141–159.

Eding H., Crooijmans R. P. M. A., Groenen M. A. M., Meuwissen T. H. E., 2002 Assessing the contribution of breeds to genetic diversity in conservation schemes. Genet. Sel. Evol. 34: 613–633.

Falconer D. S., Mackay T. F. C., 1996 *Introduction to quantitative genetics*. Harlow: Pearson Ediucation Limited.

Hartwig S., Wellmann R., Hamann H., Bennewitz J., 2014 The contribution of migrant breeds to the genetic gain of beef traits of German Vorderwald and Hinterwald cattle. J. Anim. Breed. Genet. 131: 496–503.

Hartwig S., Wellmann R., Emmerling R., Hamann H., Bennewitz J., 2015 Short communication: Importance of introgression for milk traits in the German Vorderwald and Hinterwald cattle. J. Dairy Sci. 98: 2033–2038.

Meuwissen T. H. E., 1997 Maximizing the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75: 2575–2583.

Meuwissen T. H. E., 2009 Editorial: Towards consensus on how to measure neutral genetic diversity? J. Anim. Breed. Genet. 126: 333–334.

Pfaff B., 2014 The R package cccp: Design for solving cone constrained convex programs. R Financ.: http://www.pfaffikus.de/files/conf/rif/rif2014.pdf.

Piyasatian N., Kinghorn B. P., 2003 Balancing genetic diversity, genetic merit and population viability in conservation programmes. J. Anim. Breed. Genet. 120: 137–149.

Taberlet P., Valentini A., Rezaei H. R., Naderi S., Pompanon F., *et al.*, 2008 Are cattle, sheep, and goats endangered species? Mol. Ecol. 17: 275–284.

Todesco M., Pascual M. A., Owens G. L., Ostevik K. L., Moyers B. T., *et al.*, 2016 Hybridization and extinction. Evol. Appl. 9: 892–908.

Toro M. Á., Fernández J., Caballero A., 2006 Scientific basis for policies in conservation of farm animal genetic resources. In: *Proceedings of the 8th world congress on genetics applied to livestock production:* 13–18 August 2006; Belo Horizonte.

Toro M. Á., Villanueva B., Fernández J., 2014 Genomics applied to management strategies in conservation programmes. Livest. Sci. 166: 48–53.

Wang Y., Bennewitz J., Wellmann R., 2017a Novel optimum contribution selection methods accounting for conflicting objectives in breeding programs for livestock breeds with historical migration. Genet. Sel. Evol. 49: 45.

Wang Y., Segelke D., Emmerling R., Bennewitz J., Wellmann R., 2017b Long-term impact of optimum contribution selection strategies on local livestock breeds with historical introgression at the example of German Angler cattle. G3 Genes|Genomes|Genetics 7: 4009–4018.

Wellmann R., Hartwig S., Bennewitz J., 2012 Optimum contribution selection for conserved populations with historic migration. Genet. Sel. Evol. 44: 34.

Wellmann R., Bennewitz J., Meuwissen T. H. E., 2014 A unified approach to characterize and conserve adaptive and neutral genetic diversity in subdivided populations. Genet. Res. (Camb). 96: e16.

Wellmann R., 2018 Optimum contribution selection and mate allocation for breeding: the R package *optiSel*. https://cran.r-project.org/web/packages/optiSel/optiSel.pdf.

Weng Z.-Q., Saatchi M., Schnabel R. D., Taylor J. F., Garrick D. J., 2014 Recombination locations and rates in beef cattle assessed from parent-offspring pairs. Genet. Sel. Evol. 46: 34.

Woolliams J. A., Berg P., Dagnachew B. S., Meuwissen T. H. E., 2015 Genetic contributions and their optimization. J. Anim. Breed. Genet. 132: 89–99.

# Additional Files

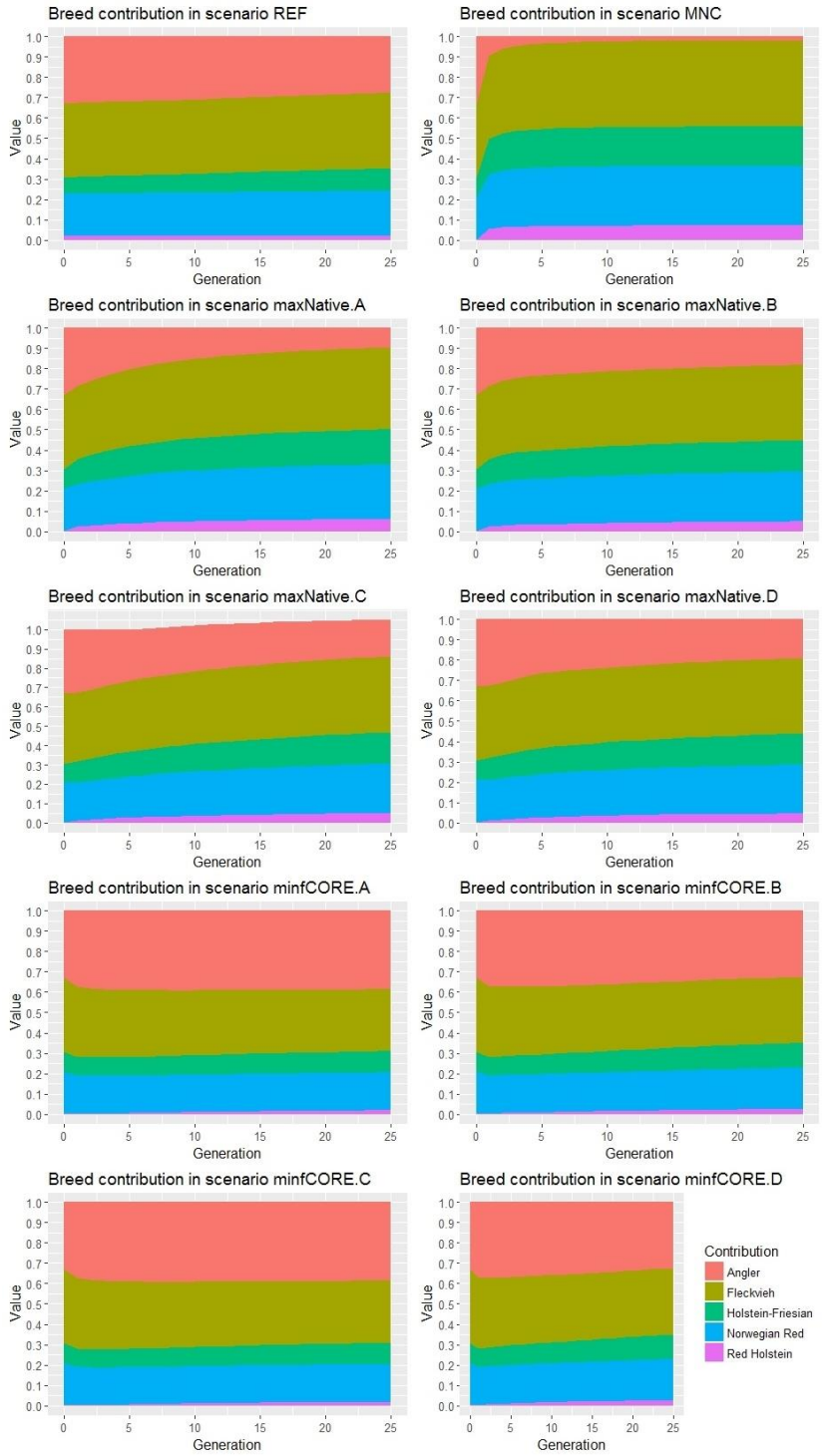



**Figure 5.S1** Breed contribution to the hypothetical multi-breed population to maximize neutral gene diversity in each scenario. **Figure 5.S1a** (Above) Scenario REF and Scenario MNC; **Figure 5.S1b (Middle)** Scenarios aiming at maximizing the native contribution of Angler cattle; **Figure 5.S1c (Bottom)** Scenarios aiming at minimizing kinship across breeds.

**File 5.S1**: SupplementTables.xlsx

**Table5.S1**. Average native contribution (NC) of Angler cattle achieved in each generation of each scenario

| | REF | | MNC | | maxNative.A | | maxNative.B | | maxNative.C | | maxNative.D | | minfCORE.A | | minfCORE.B | | minfCORE.C | | minfCORE.D | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| G0 | 0.3171 | 0.0000 | 0.3171 | 0.0000 | 0.3171 | 0.0000 | 0.3171 | 0.0000 | 0.3171 | 0.0000 | 0.3171 | 0.0000 | 0.3171 | 0.0000 | 0.3171 | 0.0000 | 0.3171 | 0.0000 | 0.3171 | 0.0000 |
| G1 | 0.3171 | 0.0004 | 0.4339 | 0.0013 | 0.3976 | 0.0014 | 0.3975 | 0.0012 | 0.3869 | 0.0011 | 0.3866 | 0.0009 | 0.3628 | 0.0005 | 0.3625 | 0.0006 | 0.3596 | 0.0003 | 0.3605 | 0.0006 |
| G2 | 0.3177 | 0.0007 | 0.4887 | 0.0060 | 0.4336 | 0.0033 | 0.4335 | 0.0018 | 0.4135 | 0.0017 | 0.4136 | 0.0017 | 0.3651 | 0.0015 | 0.3777 | 0.0014 | 0.3638 | 0.0014 | 0.3767 | 0.0027 |
| G3 | 0.3174 | 0.0006 | 0.5277 | 0.0099 | 0.4613 | 0.0045 | 0.4583 | 0.0026 | 0.4357 | 0.0035 | 0.4358 | 0.0026 | 0.3655 | 0.0015 | 0.3872 | 0.0027 | 0.3643 | 0.0019 | 0.3860 | 0.0040 |
| G4 | 0.3171 | 0.0009 | 0.5587 | 0.0107 | 0.4866 | 0.0037 | 0.4808 | 0.0043 | 0.4561 | 0.0030 | 0.4563 | 0.0022 | 0.3658 | 0.0025 | 0.3953 | 0.0037 | 0.3656 | 0.0016 | 0.3960 | 0.0042 |
| G5 | 0.3175 | 0.0010 | 0.5816 | 0.0091 | 0.5087 | 0.0034 | 0.4995 | 0.0056 | 0.4754 | 0.0030 | 0.4753 | 0.0033 | 0.3655 | 0.0032 | 0.4034 | 0.0029 | 0.3663 | 0.0025 | 0.4045 | 0.0042 |
| G6 | 0.3177 | 0.0011 | 0.5995 | 0.0084 | 0.5302 | 0.0034 | 0.5161 | 0.0060 | 0.4933 | 0.0027 | 0.4930 | 0.0045 | 0.3653 | 0.0035 | 0.4124 | 0.0034 | 0.3662 | 0.0036 | 0.4125 | 0.0042 |
| G7 | 0.3178 | 0.0015 | 0.6154 | 0.0070 | 0.5514 | 0.0032 | 0.5320 | 0.0063 | 0.5107 | 0.0021 | 0.5098 | 0.0050 | 0.3647 | 0.0057 | 0.4207 | 0.0047 | 0.3652 | 0.0035 | 0.4207 | 0.0039 |
| G8 | 0.3180 | 0.0017 | 0.6306 | 0.0073 | 0.5712 | 0.0038 | 0.5469 | 0.0064 | 0.5285 | 0.0022 | 0.5264 | 0.0054 | 0.3655 | 0.0052 | 0.4286 | 0.0055 | 0.3656 | 0.0034 | 0.4283 | 0.0046 |
| G9 | 0.3180 | 0.0019 | 0.6417 | 0.0077 | 0.5892 | 0.0049 | 0.5615 | 0.0077 | 0.5452 | 0.0019 | 0.5426 | 0.0055 | 0.3657 | 0.0051 | 0.4377 | 0.0074 | 0.3655 | 0.0037 | 0.4356 | 0.0052 |
| G10 | 0.3178 | 0.0022 | 0.6509 | 0.0076 | 0.6067 | 0.0051 | 0.5753 | 0.0079 | 0.5612 | 0.0024 | 0.5570 | 0.0062 | 0.3660 | 0.0049 | 0.4462 | 0.0085 | 0.3662 | 0.0035 | 0.4427 | 0.0054 |
| G11 | 0.3181 | 0.0026 | 0.6592 | 0.0071 | 0.6228 | 0.0065 | 0.5894 | 0.0088 | 0.5768 | 0.0025 | 0.5714 | 0.0056 | 0.3654 | 0.0051 | 0.4547 | 0.0099 | 0.3659 | 0.0043 | 0.4499 | 0.0064 |
| G12 | 0.3181 | 0.0027 | 0.6658 | 0.0067 | 0.6375 | 0.0067 | 0.6029 | 0.0081 | 0.5921 | 0.0025 | 0.5846 | 0.0057 | 0.3648 | 0.0051 | 0.4623 | 0.0102 | 0.3658 | 0.0041 | 0.4576 | 0.0068 |
| G13 | 0.3183 | 0.0030 | 0.6714 | 0.0066 | 0.6536 | 0.0079 | 0.6155 | 0.0086 | 0.6063 | 0.0025 | 0.5973 | 0.0060 | 0.3650 | 0.0056 | 0.4699 | 0.0108 | 0.3664 | 0.0042 | 0.4654 | 0.0072 |
| G14 | 0.3182 | 0.0031 | 0.6764 | 0.0068 | 0.6679 | 0.0091 | 0.6280 | 0.0093 | 0.6198 | 0.0027 | 0.6096 | 0.0059 | 0.3653 | 0.0052 | 0.4790 | 0.0107 | 0.3665 | 0.0044 | 0.4728 | 0.0076 |
| G15 | 0.3184 | 0.0031 | 0.6799 | 0.0069 | 0.6810 | 0.0089 | 0.6399 | 0.0100 | 0.6331 | 0.0020 | 0.6217 | 0.0062 | 0.3653 | 0.0053 | 0.4882 | 0.0113 | 0.3662 | 0.0041 | 0.4801 | 0.0090 |
| G16 | 0.3185 | 0.0032 | 0.6826 | 0.0063 | 0.6934 | 0.0094 | 0.6504 | 0.0107 | 0.6460 | 0.0017 | 0.6337 | 0.0057 | 0.3653 | 0.0055 | 0.4958 | 0.0129 | 0.3665 | 0.0037 | 0.4878 | 0.0098 |
| G17 | 0.3184 | 0.0034 | 0.6851 | 0.0063 | 0.7052 | 0.0100 | 0.6605 | 0.0111 | 0.6580 | 0.0023 | 0.6446 | 0.0058 | 0.3653 | 0.0055 | 0.5046 | 0.0145 | 0.3669 | 0.0035 | 0.4954 | 0.0107 |
| G18 | 0.3183 | 0.0036 | 0.6873 | 0.0065 | 0.7178 | 0.0097 | 0.6704 | 0.0103 | 0.6696 | 0.0023 | 0.6558 | 0.0058 | 0.3648 | 0.0053 | 0.5127 | 0.0156 | 0.3663 | 0.0025 | 0.5024 | 0.0110 |
| G19 | 0.3184 | 0.0039 | 0.6896 | 0.0064 | 0.7292 | 0.0098 | 0.6806 | 0.0102 | 0.6813 | 0.0026 | 0.6658 | 0.0058 | 0.3648 | 0.0058 | 0.5200 | 0.0164 | 0.3669 | 0.0026 | 0.5105 | 0.0115 |
| G20 | 0.3187 | 0.0040 | 0.6917 | 0.0061 | 0.7402 | 0.0099 | 0.6901 | 0.0100 | 0.6923 | 0.0029 | 0.6762 | 0.0056 | 0.3654 | 0.0058 | 0.5274 | 0.0174 | 0.3665 | 0.0021 | 0.5192 | 0.0111 |
| G21 | 0.3188 | 0.0042 | 0.6932 | 0.0061 | 0.7507 | 0.0096 | 0.6996 | 0.0099 | 0.7029 | 0.0031 | 0.6860 | 0.0059 | 0.3652 | 0.0057 | 0.5338 | 0.0182 | 0.3663 | 0.0023 | 0.5263 | 0.0116 |
| G22 | 0.3188 | 0.0040 | 0.6948 | 0.0062 | 0.7606 | 0.0092 | 0.7083 | 0.0096 | 0.7131 | 0.0038 | 0.6962 | 0.0056 | 0.3654 | 0.0058 | 0.5402 | 0.0192 | 0.3665 | 0.0018 | 0.5341 | 0.0125 |
| G23 | 0.3188 | 0.0044 | 0.6960 | 0.0059 | 0.7699 | 0.0091 | 0.7172 | 0.0092 | 0.7228 | 0.0042 | 0.7059 | 0.0062 | 0.3658 | 0.0063 | 0.5475 | 0.0209 | 0.3664 | 0.0023 | 0.5406 | 0.0128 |
| G24 | 0.3188 | 0.0045 | 0.6968 | 0.0057 | 0.7783 | 0.0090 | 0.7260 | 0.0090 | 0.7322 | 0.0039 | 0.7154 | 0.0060 | 0.3660 | 0.0061 | 0.5543 | 0.0221 | 0.3659 | 0.0021 | 0.5469 | 0.0134 |
| G25 | 0.3189 | 0.0048 | 0.6975 | 0.0054 | 0.7870 | 0.0092 | 0.7337 | 0.0096 | 0.7415 | 0.0040 | 0.7243 | 0.0060 | 0.3665 | 0.0065 | 0.5617 | 0.0226 | 0.3661 | 0.0021 | 0.5528 | 0.0147 |

**Table5.S2** Average kinship across breeds ($f_{CORE}$) achieved in each generation of each scenario

| | REF | | MNC | | maxNative.A | | maxNative.B | | maxNative.C | | maxNative.D | | minfCORE.A | | minfCORE.B | | minfCORE.C | | minfCORE.D | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| G0 | 0.0285 | 0.0000 | 0.0285 | 0.0000 | 0.0285 | 0.0000 | 0.0285 | 0.0000 | 0.0285 | 0.0000 | 0.0285 | 0.0000 | 0.0285 | 0.0000 | 0.0285 | 0.0000 | 0.0285 | 0.0000 | 0.0285 | 0.0000 |
| G1 | 0.0285 | 0.0000 | 0.0309 | 0.0000 | 0.0282 | 0.0000 | 0.0282 | 0.0000 | 0.0277 | 0.0000 | 0.0277 | 0.0000 | 0.0271 | 0.0000 | 0.0271 | 0.0001 | 0.0272 | 0.0000 | 0.0272 | 0.0000 |
| G2 | 0.0285 | 0.0001 | 0.0315 | 0.0001 | 0.0284 | 0.0001 | 0.0284 | 0.0000 | 0.0279 | 0.0000 | 0.0278 | 0.0001 | 0.0267 | 0.0001 | 0.0268 | 0.0001 | 0.0268 | 0.0000 | 0.0268 | 0.0001 |
| G3 | 0.0284 | 0.0001 | 0.0317 | 0.0001 | 0.0287 | 0.0000 | 0.0284 | 0.0000 | 0.0281 | 0.0001 | 0.0281 | 0.0001 | 0.0264 | 0.0001 | 0.0265 | 0.0001 | 0.0264 | 0.0000 | 0.0266 | 0.0001 |
| G4 | 0.0284 | 0.0001 | 0.0318 | 0.0001 | 0.0290 | 0.0000 | 0.0285 | 0.0000 | 0.0283 | 0.0001 | 0.0283 | 0.0001 | 0.0261 | 0.0001 | 0.0263 | 0.0001 | 0.0262 | 0.0000 | 0.0264 | 0.0001 |
| G5 | 0.0284 | 0.0001 | 0.0319 | 0.0000 | 0.0292 | 0.0000 | 0.0285 | 0.0000 | 0.0284 | 0.0001 | 0.0284 | 0.0001 | 0.0258 | 0.0001 | 0.0261 | 0.0001 | 0.0260 | 0.0000 | 0.0263 | 0.0001 |
| G6 | 0.0284 | 0.0001 | 0.0319 | 0.0000 | 0.0294 | 0.0000 | 0.0285 | 0.0000 | 0.0286 | 0.0001 | 0.0285 | 0.0000 | 0.0256 | 0.0001 | 0.0260 | 0.0002 | 0.0258 | 0.0000 | 0.0262 | 0.0001 |
| G7 | 0.0284 | 0.0001 | 0.0320 | 0.0000 | 0.0296 | 0.0000 | 0.0285 | 0.0000 | 0.0287 | 0.0001 | 0.0285 | 0.0000 | 0.0255 | 0.0001 | 0.0259 | 0.0002 | 0.0256 | 0.0000 | 0.0261 | 0.0001 |
| G8 | 0.0284 | 0.0000 | 0.0320 | 0.0000 | 0.0297 | 0.0000 | 0.0284 | 0.0000 | 0.0289 | 0.0001 | 0.0284 | 0.0000 | 0.0253 | 0.0000 | 0.0258 | 0.0002 | 0.0254 | 0.0000 | 0.0260 | 0.0001 |
| G9 | 0.0284 | 0.0001 | 0.0320 | 0.0000 | 0.0298 | 0.0000 | 0.0285 | 0.0000 | 0.0290 | 0.0001 | 0.0284 | 0.0000 | 0.0251 | 0.0000 | 0.0257 | 0.0002 | 0.0253 | 0.0000 | 0.0259 | 0.0001 |
| G10 | 0.0285 | 0.0001 | 0.0321 | 0.0000 | 0.0299 | 0.0001 | 0.0285 | 0.0000 | 0.0291 | 0.0001 | 0.0284 | 0.0000 | 0.0250 | 0.0000 | 0.0256 | 0.0002 | 0.0252 | 0.0000 | 0.0258 | 0.0001 |
| G11 | 0.0285 | 0.0000 | 0.0321 | 0.0000 | 0.0300 | 0.0001 | 0.0285 | 0.0000 | 0.0292 | 0.0000 | 0.0285 | 0.0000 | 0.0249 | 0.0000 | 0.0255 | 0.0002 | 0.0250 | 0.0000 | 0.0257 | 0.0001 |
| G12 | 0.0285 | 0.0001 | 0.0321 | 0.0000 | 0.0301 | 0.0001 | 0.0285 | 0.0000 | 0.0293 | 0.0001 | 0.0285 | 0.0000 | 0.0248 | 0.0000 | 0.0255 | 0.0003 | 0.0249 | 0.0000 | 0.0256 | 0.0001 |
| G13 | 0.0285 | 0.0001 | 0.0321 | 0.0000 | 0.0302 | 0.0001 | 0.0285 | 0.0000 | 0.0294 | 0.0001 | 0.0284 | 0.0000 | 0.0247 | 0.0000 | 0.0254 | 0.0003 | 0.0248 | 0.0001 | 0.0256 | 0.0001 |
| G14 | 0.0286 | 0.0001 | 0.0321 | 0.0000 | 0.0302 | 0.0001 | 0.0285 | 0.0000 | 0.0295 | 0.0001 | 0.0284 | 0.0000 | 0.0246 | 0.0000 | 0.0254 | 0.0002 | 0.0247 | 0.0001 | 0.0255 | 0.0001 |
| G15 | 0.0286 | 0.0001 | 0.0321 | 0.0000 | 0.0303 | 0.0001 | 0.0285 | 0.0000 | 0.0295 | 0.0000 | 0.0284 | 0.0000 | 0.0245 | 0.0000 | 0.0254 | 0.0003 | 0.0246 | 0.0000 | 0.0255 | 0.0002 |
| G16 | 0.0286 | 0.0001 | 0.0321 | 0.0000 | 0.0304 | 0.0001 | 0.0285 | 0.0000 | 0.0296 | 0.0000 | 0.0285 | 0.0000 | 0.0244 | 0.0000 | 0.0253 | 0.0003 | 0.0245 | 0.0001 | 0.0254 | 0.0002 |
| G17 | 0.0286 | 0.0001 | 0.0321 | 0.0000 | 0.0304 | 0.0001 | 0.0285 | 0.0000 | 0.0297 | 0.0000 | 0.0285 | 0.0000 | 0.0243 | 0.0000 | 0.0253 | 0.0003 | 0.0245 | 0.0001 | 0.0254 | 0.0002 |
| G18 | 0.0286 | 0.0001 | 0.0321 | 0.0000 | 0.0305 | 0.0001 | 0.0285 | 0.0000 | 0.0298 | 0.0000 | 0.0285 | 0.0000 | 0.0242 | 0.0000 | 0.0253 | 0.0004 | 0.0244 | 0.0001 | 0.0254 | 0.0002 |
| G19 | 0.0287 | 0.0001 | 0.0321 | 0.0000 | 0.0305 | 0.0001 | 0.0285 | 0.0000 | 0.0298 | 0.0000 | 0.0284 | 0.0000 | 0.0241 | 0.0000 | 0.0253 | 0.0004 | 0.0243 | 0.0001 | 0.0254 | 0.0002 |
| G20 | 0.0287 | 0.0001 | 0.0321 | 0.0000 | 0.0306 | 0.0001 | 0.0285 | 0.0000 | 0.0299 | 0.0001 | 0.0285 | 0.0000 | 0.0240 | 0.0000 | 0.0253 | 0.0004 | 0.0242 | 0.0001 | 0.0254 | 0.0002 |
| G21 | 0.0287 | 0.0001 | 0.0321 | 0.0000 | 0.0306 | 0.0001 | 0.0285 | 0.0000 | 0.0299 | 0.0001 | 0.0285 | 0.0000 | 0.0240 | 0.0000 | 0.0252 | 0.0004 | 0.0242 | 0.0001 | 0.0253 | 0.0002 |
| G22 | 0.0287 | 0.0001 | 0.0321 | 0.0000 | 0.0307 | 0.0001 | 0.0285 | 0.0000 | 0.0300 | 0.0001 | 0.0285 | 0.0000 | 0.0239 | 0.0000 | 0.0252 | 0.0004 | 0.0241 | 0.0001 | 0.0254 | 0.0003 |
| G23 | 0.0288 | 0.0001 | 0.0322 | 0.0000 | 0.0307 | 0.0001 | 0.0285 | 0.0000 | 0.0300 | 0.0001 | 0.0284 | 0.0000 | 0.0238 | 0.0000 | 0.0252 | 0.0004 | 0.0240 | 0.0001 | 0.0253 | 0.0003 |
| G24 | 0.0288 | 0.0001 | 0.0322 | 0.0000 | 0.0307 | 0.0001 | 0.0285 | 0.0000 | 0.0301 | 0.0001 | 0.0285 | 0.0000 | 0.0238 | 0.0000 | 0.0252 | 0.0004 | 0.0240 | 0.0001 | 0.0253 | 0.0003 |
| G25 | 0.0288 | 0.0001 | 0.0322 | 0.0000 | 0.0308 | 0.0001 | 0.0285 | 0.0000 | 0.0301 | 0.0001 | 0.0285 | 0.0000 | 0.0237 | 0.0000 | 0.0251 | 0.0004 | 0.0239 | 0.0001 | 0.0253 | 0.0003 |

**Table5.S3** Average kinship of Angler cattle ($f_{SEG}$) achieved in each generation of each scenario

| | REF | | MNC | | maxNative.A | | maxNative.B | | maxNative.C | | maxNative.D | | minfCORE.A | | minfCORE.B | | minfCORE.C | | minfCORE.D | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| G0 | 0.0546 | 0.0000 | 0.0546 | 0.0000 | 0.0546 | 0.0000 | 0.0546 | 0.0000 | 0.0546 | 0.0000 | 0.0546 | 0.0000 | 0.0546 | 0.0000 | 0.0546 | 0.0000 | 0.0546 | 0.0000 | 0.0546 | 0.0000 |
| G1 | 0.0550 | 0.0002 | 0.1707 | 0.0002 | 0.0608 | 0.0002 | 0.0607 | 0.0002 | 0.0540 | 0.0001 | 0.0539 | 0.0002 | 0.0491 | 0.0001 | 0.0491 | 0.0003 | 0.0491 | 0.0002 | 0.0492 | 0.0001 |
| G2 | 0.0552 | 0.0003 | 0.2711 | 0.0175 | 0.0666 | 0.0002 | 0.0665 | 0.0002 | 0.0564 | 0.0002 | 0.0563 | 0.0005 | 0.0481 | 0.0003 | 0.0488 | 0.0003 | 0.0482 | 0.0001 | 0.0489 | 0.0004 |
| G3 | 0.0554 | 0.0003 | 0.3574 | 0.0220 | 0.0727 | 0.0002 | 0.0708 | 0.0004 | 0.0592 | 0.0004 | 0.0593 | 0.0006 | 0.0475 | 0.0003 | 0.0487 | 0.0004 | 0.0475 | 0.0001 | 0.0489 | 0.0004 |
| G4 | 0.0555 | 0.0003 | 0.4286 | 0.0385 | 0.0786 | 0.0003 | 0.0735 | 0.0004 | 0.0621 | 0.0007 | 0.0620 | 0.0007 | 0.0472 | 0.0003 | 0.0487 | 0.0005 | 0.0472 | 0.0002 | 0.0491 | 0.0002 |
| G5 | 0.0558 | 0.0003 | 0.4851 | 0.0421 | 0.0846 | 0.0006 | 0.0756 | 0.0008 | 0.0648 | 0.0007 | 0.0649 | 0.0007 | 0.0469 | 0.0004 | 0.0488 | 0.0006 | 0.0470 | 0.0002 | 0.0493 | 0.0003 |
| G6 | 0.0560 | 0.0003 | 0.5375 | 0.0413 | 0.0902 | 0.0006 | 0.0778 | 0.0009 | 0.0676 | 0.0004 | 0.0670 | 0.0008 | 0.0467 | 0.0003 | 0.0490 | 0.0006 | 0.0467 | 0.0002 | 0.0495 | 0.0004 |
| G7 | 0.0563 | 0.0002 | 0.5917 | 0.0331 | 0.0960 | 0.0007 | 0.0801 | 0.0009 | 0.0706 | 0.0004 | 0.0687 | 0.0005 | 0.0466 | 0.0003 | 0.0493 | 0.0008 | 0.0466 | 0.0001 | 0.0498 | 0.0004 |
| G8 | 0.0565 | 0.0001 | 0.6344 | 0.0271 | 0.1018 | 0.0008 | 0.0823 | 0.0009 | 0.0734 | 0.0005 | 0.0706 | 0.0008 | 0.0466 | 0.0003 | 0.0496 | 0.0008 | 0.0465 | 0.0001 | 0.0500 | 0.0006 |
| G9 | 0.0569 | 0.0002 | 0.6781 | 0.0239 | 0.1075 | 0.0011 | 0.0845 | 0.0006 | 0.0763 | 0.0006 | 0.0725 | 0.0009 | 0.0466 | 0.0003 | 0.0500 | 0.0011 | 0.0465 | 0.0002 | 0.0504 | 0.0007 |
| G10 | 0.0572 | 0.0003 | 0.7173 | 0.0196 | 0.1131 | 0.0012 | 0.0864 | 0.0008 | 0.0791 | 0.0009 | 0.0746 | 0.0009 | 0.0466 | 0.0003 | 0.0505 | 0.0012 | 0.0466 | 0.0002 | 0.0507 | 0.0007 |
| G11 | 0.0576 | 0.0002 | 0.7499 | 0.0196 | 0.1185 | 0.0013 | 0.0883 | 0.0009 | 0.0818 | 0.0007 | 0.0765 | 0.0009 | 0.0466 | 0.0002 | 0.0509 | 0.0014 | 0.0466 | 0.0002 | 0.0509 | 0.0008 |
| G12 | 0.0580 | 0.0004 | 0.7737 | 0.0108 | 0.1239 | 0.0014 | 0.0902 | 0.0010 | 0.0850 | 0.0009 | 0.0785 | 0.0012 | 0.0468 | 0.0002 | 0.0514 | 0.0015 | 0.0466 | 0.0003 | 0.0514 | 0.0008 |
| G13 | 0.0584 | 0.0003 | 0.8025 | 0.0109 | 0.1292 | 0.0015 | 0.0919 | 0.0010 | 0.0880 | 0.0009 | 0.0804 | 0.0012 | 0.0468 | 0.0002 | 0.0518 | 0.0015 | 0.0467 | 0.0004 | 0.0518 | 0.0008 |
| G14 | 0.0587 | 0.0004 | 0.8218 | 0.0073 | 0.1345 | 0.0014 | 0.0938 | 0.0011 | 0.0907 | 0.0007 | 0.0824 | 0.0013 | 0.0468 | 0.0002 | 0.0524 | 0.0014 | 0.0468 | 0.0004 | 0.0522 | 0.0008 |
| G15 | 0.0590 | 0.0004 | 0.8428 | 0.0072 | 0.1398 | 0.0013 | 0.0957 | 0.0013 | 0.0939 | 0.0006 | 0.0842 | 0.0012 | 0.0469 | 0.0002 | 0.0531 | 0.0017 | 0.0468 | 0.0004 | 0.0527 | 0.0010 |
| G16 | 0.0594 | 0.0005 | 0.8569 | 0.0102 | 0.1449 | 0.0016 | 0.0975 | 0.0014 | 0.0972 | 0.0007 | 0.0864 | 0.0012 | 0.0469 | 0.0003 | 0.0537 | 0.0019 | 0.0469 | 0.0003 | 0.0532 | 0.0014 |
| G17 | 0.0597 | 0.0006 | 0.8697 | 0.0127 | 0.1501 | 0.0018 | 0.0993 | 0.0013 | 0.1002 | 0.0009 | 0.0886 | 0.0010 | 0.0470 | 0.0003 | 0.0545 | 0.0021 | 0.0469 | 0.0005 | 0.0537 | 0.0016 |
| G18 | 0.0600 | 0.0007 | 0.8823 | 0.0122 | 0.1552 | 0.0017 | 0.1008 | 0.0009 | 0.1030 | 0.0011 | 0.0903 | 0.0010 | 0.0471 | 0.0003 | 0.0551 | 0.0025 | 0.0471 | 0.0005 | 0.0542 | 0.0015 |
| G19 | 0.0605 | 0.0007 | 0.8899 | 0.0222 | 0.1605 | 0.0018 | 0.1024 | 0.0011 | 0.1063 | 0.0012 | 0.0920 | 0.0011 | 0.0471 | 0.0003 | 0.0556 | 0.0026 | 0.0471 | 0.0005 | 0.0548 | 0.0017 |
| G20 | 0.0609 | 0.0008 | 0.9010 | 0.0173 | 0.1656 | 0.0017 | 0.1039 | 0.0009 | 0.1094 | 0.0014 | 0.0941 | 0.0011 | 0.0472 | 0.0003 | 0.0562 | 0.0027 | 0.0472 | 0.0005 | 0.0555 | 0.0017 |
| G21 | 0.0613 | 0.0008 | 0.9074 | 0.0154 | 0.1706 | 0.0020 | 0.1058 | 0.0006 | 0.1125 | 0.0018 | 0.0958 | 0.0011 | 0.0473 | 0.0003 | 0.0566 | 0.0028 | 0.0473 | 0.0005 | 0.0560 | 0.0017 |
| G22 | 0.0616 | 0.0009 | 0.9199 | 0.0128 | 0.1756 | 0.0022 | 0.1074 | 0.0006 | 0.1156 | 0.0015 | 0.0974 | 0.0011 | 0.0475 | 0.0003 | 0.0570 | 0.0029 | 0.0474 | 0.0006 | 0.0568 | 0.0019 |
| G23 | 0.0619 | 0.0010 | 0.9288 | 0.0105 | 0.1806 | 0.0023 | 0.1091 | 0.0008 | 0.1187 | 0.0014 | 0.0992 | 0.0013 | 0.0476 | 0.0004 | 0.0576 | 0.0033 | 0.0475 | 0.0005 | 0.0571 | 0.0019 |
| G24 | 0.0622 | 0.0011 | 0.9363 | 0.0089 | 0.1854 | 0.0025 | 0.1107 | 0.0010 | 0.1221 | 0.0014 | 0.1012 | 0.0011 | 0.0477 | 0.0003 | 0.0581 | 0.0035 | 0.0476 | 0.0005 | 0.0577 | 0.0020 |
| G25 | 0.0625 | 0.0010 | 0.9422 | 0.0104 | 0.1904 | 0.0025 | 0.1122 | 0.0010 | 0.1254 | 0.0016 | 0.1025 | 0.0011 | 0.0478 | 0.0003 | 0.0586 | 0.0037 | 0.0477 | 0.0006 | 0.0581 | 0.0022 |

**Table5.S4** Average kinship at native alleles of Angler cattle ($f_{SEG|N}$) achieved in each generation of each scenario

| | REF | | MNC | | maxNative.A | | maxNative.B | | maxNative.C | | maxNative.D | | minfCORE.A | | minfCORE.B | | minfCORE.C | | minfCORE.D | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| G0 | 0.0763 | 0.0000 | 0.0763 | 0.0000 | 0.0763 | 0.0000 | 0.0763 | 0.0000 | 0.0763 | 0.0000 | 0.0763 | 0.0000 | 0.0763 | 0.0000 | 0.0763 | 0.0000 | 0.0763 | 0.0000 | 0.0763 | 0.0000 |
| G1 | 0.0761 | 0.0003 | 0.3211 | 0.0008 | 0.1000 | 0.0004 | 0.0998 | 0.0005 | 0.0810 | 0.0002 | 0.0809 | 0.0002 | 0.0718 | 0.0001 | 0.0719 | 0.0004 | 0.0715 | 0.0003 | 0.0716 | 0.0003 |
| G2 | 0.0762 | 0.0006 | 0.4485 | 0.0299 | 0.1138 | 0.0008 | 0.1137 | 0.0007 | 0.0854 | 0.0001 | 0.0854 | 0.0004 | 0.0731 | 0.0005 | 0.0750 | 0.0004 | 0.0724 | 0.0005 | 0.0747 | 0.0009 |
| G3 | 0.0766 | 0.0007 | 0.5579 | 0.0400 | 0.1265 | 0.0011 | 0.1232 | 0.0011 | 0.0896 | 0.0003 | 0.0899 | 0.0004 | 0.0739 | 0.0007 | 0.0762 | 0.0005 | 0.0731 | 0.0004 | 0.0764 | 0.0010 |
| G4 | 0.0762 | 0.0008 | 0.6264 | 0.0556 | 0.1380 | 0.0022 | 0.1282 | 0.0009 | 0.0941 | 0.0005 | 0.0944 | 0.0003 | 0.0742 | 0.0007 | 0.0772 | 0.0006 | 0.0737 | 0.0008 | 0.0776 | 0.0010 |
| G5 | 0.0762 | 0.0009 | 0.6684 | 0.0461 | 0.1488 | 0.0029 | 0.1315 | 0.0011 | 0.0984 | 0.0005 | 0.0988 | 0.0005 | 0.0743 | 0.0006 | 0.0784 | 0.0009 | 0.0741 | 0.0008 | 0.0785 | 0.0010 |
| G6 | 0.0766 | 0.0009 | 0.7107 | 0.0406 | 0.1590 | 0.0033 | 0.1345 | 0.0009 | 0.1025 | 0.0008 | 0.1031 | 0.0006 | 0.0745 | 0.0005 | 0.0795 | 0.0011 | 0.0741 | 0.0008 | 0.0792 | 0.0012 |
| G7 | 0.0770 | 0.0010 | 0.7457 | 0.0281 | 0.1685 | 0.0022 | 0.1382 | 0.0014 | 0.1064 | 0.0008 | 0.1072 | 0.0003 | 0.0748 | 0.0003 | 0.0803 | 0.0014 | 0.0744 | 0.0006 | 0.0801 | 0.0012 |
| G8 | 0.0772 | 0.0010 | 0.7720 | 0.0247 | 0.1782 | 0.0037 | 0.1415 | 0.0011 | 0.1104 | 0.0009 | 0.1111 | 0.0006 | 0.0749 | 0.0006 | 0.0808 | 0.0016 | 0.0747 | 0.0007 | 0.0809 | 0.0016 |
| G9 | 0.0777 | 0.0010 | 0.8090 | 0.0172 | 0.1867 | 0.0037 | 0.1448 | 0.0014 | 0.1146 | 0.0012 | 0.1150 | 0.0004 | 0.0751 | 0.0007 | 0.0815 | 0.0021 | 0.0749 | 0.0005 | 0.0817 | 0.0019 |
| G10 | 0.0780 | 0.0014 | 0.8346 | 0.0151 | 0.1955 | 0.0039 | 0.1476 | 0.0021 | 0.1187 | 0.0014 | 0.1191 | 0.0007 | 0.0750 | 0.0009 | 0.0823 | 0.0021 | 0.0749 | 0.0005 | 0.0824 | 0.0020 |
| G11 | 0.0787 | 0.0015 | 0.8574 | 0.0144 | 0.2039 | 0.0036 | 0.1498 | 0.0025 | 0.1225 | 0.0013 | 0.1226 | 0.0007 | 0.0750 | 0.0007 | 0.0829 | 0.0027 | 0.0750 | 0.0005 | 0.0827 | 0.0021 |
| G12 | 0.0794 | 0.0019 | 0.8741 | 0.0123 | 0.2117 | 0.0028 | 0.1518 | 0.0016 | 0.1263 | 0.0012 | 0.1263 | 0.0010 | 0.0751 | 0.0007 | 0.0834 | 0.0028 | 0.0748 | 0.0006 | 0.0835 | 0.0022 |
| G13 | 0.0801 | 0.0020 | 0.8898 | 0.0107 | 0.2194 | 0.0023 | 0.1538 | 0.0024 | 0.1302 | 0.0014 | 0.1295 | 0.0013 | 0.0750 | 0.0006 | 0.0839 | 0.0029 | 0.0748 | 0.0009 | 0.0840 | 0.0020 |
| G14 | 0.0808 | 0.0023 | 0.9018 | 0.0122 | 0.2266 | 0.0021 | 0.1558 | 0.0026 | 0.1338 | 0.0015 | 0.1327 | 0.0013 | 0.0750 | 0.0006 | 0.0846 | 0.0031 | 0.0747 | 0.0011 | 0.0845 | 0.0021 |
| G15 | 0.0814 | 0.0025 | 0.9127 | 0.0104 | 0.2340 | 0.0023 | 0.1581 | 0.0028 | 0.1375 | 0.0016 | 0.1356 | 0.0010 | 0.0747 | 0.0006 | 0.0857 | 0.0036 | 0.0746 | 0.0010 | 0.0849 | 0.0024 |
| G16 | 0.0821 | 0.0028 | 0.9178 | 0.0132 | 0.2403 | 0.0021 | 0.1600 | 0.0033 | 0.1411 | 0.0016 | 0.1387 | 0.0013 | 0.0744 | 0.0005 | 0.0863 | 0.0037 | 0.0745 | 0.0009 | 0.0856 | 0.0029 |
| G17 | 0.0828 | 0.0031 | 0.9253 | 0.0161 | 0.2474 | 0.0025 | 0.1618 | 0.0034 | 0.1446 | 0.0016 | 0.1416 | 0.0011 | 0.0743 | 0.0005 | 0.0873 | 0.0041 | 0.0742 | 0.0011 | 0.0861 | 0.0031 |
| G18 | 0.0835 | 0.0033 | 0.9322 | 0.0132 | 0.2533 | 0.0031 | 0.1632 | 0.0027 | 0.1481 | 0.0014 | 0.1438 | 0.0010 | 0.0740 | 0.0006 | 0.0880 | 0.0046 | 0.0741 | 0.0012 | 0.0864 | 0.0031 |
| G19 | 0.0844 | 0.0034 | 0.9391 | 0.0173 | 0.2587 | 0.0036 | 0.1648 | 0.0030 | 0.1517 | 0.0014 | 0.1460 | 0.0015 | 0.0738 | 0.0007 | 0.0883 | 0.0048 | 0.0739 | 0.0011 | 0.0871 | 0.0031 |
| G20 | 0.0850 | 0.0037 | 0.9444 | 0.0154 | 0.2650 | 0.0030 | 0.1657 | 0.0027 | 0.1550 | 0.0015 | 0.1486 | 0.0015 | 0.0736 | 0.0006 | 0.0888 | 0.0050 | 0.0738 | 0.0010 | 0.0878 | 0.0033 |
| G21 | 0.0858 | 0.0036 | 0.9463 | 0.0137 | 0.2712 | 0.0048 | 0.1676 | 0.0024 | 0.1582 | 0.0017 | 0.1505 | 0.0013 | 0.0735 | 0.0006 | 0.0890 | 0.0051 | 0.0736 | 0.0010 | 0.0883 | 0.0033 |
| G22 | 0.0864 | 0.0037 | 0.9545 | 0.0117 | 0.2773 | 0.0050 | 0.1688 | 0.0026 | 0.1618 | 0.0018 | 0.1527 | 0.0012 | 0.0734 | 0.0007 | 0.0891 | 0.0054 | 0.0735 | 0.0010 | 0.0888 | 0.0033 |
| G23 | 0.0871 | 0.0040 | 0.9586 | 0.0098 | 0.2824 | 0.0052 | 0.1702 | 0.0023 | 0.1650 | 0.0018 | 0.1551 | 0.0017 | 0.0732 | 0.0007 | 0.0898 | 0.0059 | 0.0733 | 0.0009 | 0.0889 | 0.0032 |
| G24 | 0.0877 | 0.0042 | 0.9621 | 0.0078 | 0.2872 | 0.0064 | 0.1718 | 0.0021 | 0.1682 | 0.0017 | 0.1573 | 0.0017 | 0.0730 | 0.0007 | 0.0901 | 0.0061 | 0.0731 | 0.0009 | 0.0894 | 0.0034 |
| G25 | 0.0882 | 0.0042 | 0.9660 | 0.0071 | 0.2932 | 0.0056 | 0.1728 | 0.0022 | 0.1714 | 0.0017 | 0.1584 | 0.0018 | 0.0728 | 0.0007 | 0.0904 | 0.0064 | 0.0730 | 0.0010 | 0.0895 | 0.0037 |

**Table5.S5** Optimum contribution of Angler cattle to a hypothetical multi-breed population to achieve maximum genetic diversity

| | REF | | MNC | | maxNative.A | | maxNative.B | | maxNative.C | | maxNative.D | | minfCORE.A | | minfCORE.B | | minfCORE.C | | minfCORE.D | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| G0 | 0.3273 | 0.0000 | 0.3301 | 0.0000 | 0.3301 | 0.0000 | 0.3301 | 0.0000 | 0.3301 | 0.0000 | 0.3301 | 0.0000 | 0.3301 | 0.0000 | 0.3301 | 0.0000 | 0.3301 | 0.0000 | 0.3301 | 0.0000 |
| G1 | 0.3250 | 0.0016 | 0.0970 | 0.0001 | 0.2842 | 0.0011 | 0.2845 | 0.0011 | 0.3264 | 0.0010 | 0.3268 | 0.0012 | 0.3736 | 0.0013 | 0.3740 | 0.0024 | 0.3737 | 0.0014 | 0.3734 | 0.0010 |
| G2 | 0.3235 | 0.0028 | 0.0617 | 0.0041 | 0.2588 | 0.0008 | 0.2597 | 0.0007 | 0.3094 | 0.0013 | 0.3099 | 0.0036 | 0.3807 | 0.0024 | 0.3735 | 0.0022 | 0.3816 | 0.0015 | 0.3723 | 0.0030 |
| G3 | 0.3222 | 0.0026 | 0.0475 | 0.0036 | 0.2368 | 0.0007 | 0.2464 | 0.0014 | 0.2929 | 0.0024 | 0.2924 | 0.0039 | 0.3851 | 0.0025 | 0.3735 | 0.0031 | 0.3862 | 0.0012 | 0.3723 | 0.0034 |
| G4 | 0.3211 | 0.0025 | 0.0396 | 0.0031 | 0.2186 | 0.0004 | 0.2394 | 0.0009 | 0.2782 | 0.0040 | 0.2784 | 0.0041 | 0.3876 | 0.0024 | 0.3734 | 0.0045 | 0.3884 | 0.0017 | 0.3702 | 0.0022 |
| G5 | 0.3192 | 0.0029 | 0.0353 | 0.0020 | 0.2029 | 0.0009 | 0.2343 | 0.0020 | 0.2658 | 0.0039 | 0.2653 | 0.0031 | 0.3901 | 0.0031 | 0.3724 | 0.0052 | 0.3896 | 0.0020 | 0.3687 | 0.0023 |
| G6 | 0.3176 | 0.0028 | 0.0321 | 0.0016 | 0.1904 | 0.0010 | 0.2297 | 0.0021 | 0.2537 | 0.0020 | 0.2578 | 0.0027 | 0.3913 | 0.0027 | 0.3709 | 0.0055 | 0.3916 | 0.0015 | 0.3674 | 0.0029 |
| G7 | 0.3159 | 0.0021 | 0.0294 | 0.0012 | 0.1787 | 0.0007 | 0.2250 | 0.0021 | 0.2424 | 0.0024 | 0.2524 | 0.0017 | 0.3916 | 0.0025 | 0.3690 | 0.0064 | 0.3920 | 0.0011 | 0.3656 | 0.0032 |
| G8 | 0.3143 | 0.0013 | 0.0276 | 0.0011 | 0.1684 | 0.0009 | 0.2208 | 0.0017 | 0.2330 | 0.0025 | 0.2471 | 0.0025 | 0.3915 | 0.0024 | 0.3676 | 0.0066 | 0.3922 | 0.0008 | 0.3637 | 0.0045 |
| G9 | 0.3120 | 0.0017 | 0.0258 | 0.0010 | 0.1597 | 0.0006 | 0.2168 | 0.0010 | 0.2240 | 0.0030 | 0.2420 | 0.0026 | 0.3919 | 0.0023 | 0.3652 | 0.0085 | 0.3920 | 0.0016 | 0.3615 | 0.0052 |
| G10 | 0.3098 | 0.0021 | 0.0246 | 0.0011 | 0.1524 | 0.0011 | 0.2135 | 0.0012 | 0.2155 | 0.0031 | 0.2370 | 0.0022 | 0.3918 | 0.0024 | 0.3621 | 0.0085 | 0.3918 | 0.0016 | 0.3599 | 0.0052 |
| G11 | 0.3071 | 0.0019 | 0.0237 | 0.0011 | 0.1459 | 0.0015 | 0.2105 | 0.0016 | 0.2085 | 0.0020 | 0.2324 | 0.0023 | 0.3917 | 0.0022 | 0.3595 | 0.0099 | 0.3914 | 0.0022 | 0.3589 | 0.0053 |
| G12 | 0.3045 | 0.0031 | 0.0230 | 0.0006 | 0.1401 | 0.0020 | 0.2076 | 0.0015 | 0.2005 | 0.0027 | 0.2283 | 0.0025 | 0.3908 | 0.0019 | 0.3571 | 0.0103 | 0.3915 | 0.0025 | 0.3565 | 0.0055 |
| G13 | 0.3019 | 0.0023 | 0.0221 | 0.0006 | 0.1351 | 0.0023 | 0.2052 | 0.0016 | 0.1939 | 0.0025 | 0.2246 | 0.0023 | 0.3908 | 0.0020 | 0.3554 | 0.0099 | 0.3907 | 0.0031 | 0.3542 | 0.0054 |
| G14 | 0.2999 | 0.0027 | 0.0217 | 0.0004 | 0.1303 | 0.0024 | 0.2024 | 0.0016 | 0.1884 | 0.0019 | 0.2207 | 0.0025 | 0.3903 | 0.0019 | 0.3524 | 0.0096 | 0.3902 | 0.0035 | 0.3519 | 0.0054 |
| G15 | 0.2980 | 0.0027 | 0.0212 | 0.0005 | 0.1256 | 0.0027 | 0.2000 | 0.0017 | 0.1822 | 0.0011 | 0.2175 | 0.0018 | 0.3898 | 0.0019 | 0.3486 | 0.0108 | 0.3899 | 0.0033 | 0.3497 | 0.0065 |
| G16 | 0.2958 | 0.0032 | 0.0209 | 0.0006 | 0.1212 | 0.0025 | 0.1977 | 0.0017 | 0.1759 | 0.0012 | 0.2137 | 0.0018 | 0.3899 | 0.0020 | 0.3459 | 0.0116 | 0.3898 | 0.0027 | 0.3472 | 0.0084 |
| G17 | 0.2937 | 0.0034 | 0.0205 | 0.0005 | 0.1177 | 0.0023 | 0.1955 | 0.0015 | 0.1711 | 0.0015 | 0.2103 | 0.0015 | 0.3899 | 0.0021 | 0.3421 | 0.0124 | 0.3897 | 0.0036 | 0.3448 | 0.0090 |
| G18 | 0.2916 | 0.0041 | 0.0203 | 0.0005 | 0.1142 | 0.0022 | 0.1936 | 0.0010 | 0.1667 | 0.0019 | 0.2076 | 0.0015 | 0.3892 | 0.0022 | 0.3390 | 0.0143 | 0.3890 | 0.0040 | 0.3424 | 0.0087 |
| G19 | 0.2888 | 0.0041 | 0.0200 | 0.0004 | 0.1111 | 0.0021 | 0.1918 | 0.0012 | 0.1619 | 0.0020 | 0.2053 | 0.0016 | 0.3891 | 0.0022 | 0.3372 | 0.0148 | 0.3886 | 0.0039 | 0.3397 | 0.0094 |
| G20 | 0.2868 | 0.0045 | 0.0197 | 0.0006 | 0.1082 | 0.0020 | 0.1902 | 0.0010 | 0.1576 | 0.0026 | 0.2023 | 0.0014 | 0.3887 | 0.0022 | 0.3349 | 0.0148 | 0.3881 | 0.0038 | 0.3367 | 0.0092 |
| G21 | 0.2844 | 0.0046 | 0.0197 | 0.0007 | 0.1054 | 0.0023 | 0.1881 | 0.0006 | 0.1535 | 0.0029 | 0.2000 | 0.0014 | 0.3883 | 0.0018 | 0.3333 | 0.0151 | 0.3876 | 0.0038 | 0.3345 | 0.0091 |
| G22 | 0.2825 | 0.0048 | 0.0194 | 0.0007 | 0.1024 | 0.0025 | 0.1863 | 0.0007 | 0.1496 | 0.0024 | 0.1979 | 0.0013 | 0.3877 | 0.0024 | 0.3321 | 0.0156 | 0.3875 | 0.0041 | 0.3310 | 0.0101 |
| G23 | 0.2806 | 0.0052 | 0.0193 | 0.0008 | 0.0997 | 0.0028 | 0.1847 | 0.0007 | 0.1461 | 0.0020 | 0.1957 | 0.0015 | 0.3870 | 0.0025 | 0.3297 | 0.0169 | 0.3869 | 0.0037 | 0.3300 | 0.0097 |
| G24 | 0.2791 | 0.0059 | 0.0191 | 0.0008 | 0.0974 | 0.0028 | 0.1831 | 0.0010 | 0.1424 | 0.0018 | 0.1933 | 0.0011 | 0.3866 | 0.0023 | 0.3280 | 0.0177 | 0.3862 | 0.0035 | 0.3277 | 0.0101 |
| G25 | 0.2773 | 0.0058 | 0.0190 | 0.0007 | 0.0952 | 0.0029 | 0.1817 | 0.0011 | 0.1387 | 0.0025 | 0.1918 | 0.0012 | 0.3863 | 0.0022 | 0.3262 | 0.0182 | 0.3858 | 0.0038 | 0.3262 | 0.0109 |

**Table5.S6** Optimum contribution of Fleckvieh cattle to a hypothetical multi-breed population to achieve maximum genetic diversity

| | REF | | MNC | | maxNative.A | | maxNative.B | | maxNative.C | | maxNative.D | | minfCORE.A | | minfCORE.B | | minfCORE.C | | minfCORE.D | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| G0 | 0.3642 | 0.0000 | 0.3642 | 0.0000 | 0.3642 | 0.0000 | 0.3642 | 0.0000 | 0.3642 | 0.0000 | 0.3642 | 0.0000 | 0.3642 | 0.0000 | 0.3642 | 0.0000 | 0.3642 | 0.0000 | 0.3642 | 0.0000 |
| G1 | 0.3644 | 0.0004 | 0.4044 | 0.0001 | 0.3615 | 0.0002 | 0.3614 | 0.0003 | 0.3534 | 0.0002 | 0.3533 | 0.0004 | 0.3454 | 0.0004 | 0.3452 | 0.0008 | 0.3459 | 0.0007 | 0.3459 | 0.0006 |
| G2 | 0.3643 | 0.0008 | 0.4131 | 0.0012 | 0.3657 | 0.0007 | 0.3651 | 0.0002 | 0.3565 | 0.0003 | 0.3561 | 0.0010 | 0.3396 | 0.0008 | 0.3407 | 0.0012 | 0.3403 | 0.0007 | 0.3413 | 0.0012 |
| G3 | 0.3642 | 0.0008 | 0.4163 | 0.0011 | 0.3700 | 0.0006 | 0.3664 | 0.0006 | 0.3598 | 0.0007 | 0.3598 | 0.0014 | 0.3355 | 0.0010 | 0.3376 | 0.0016 | 0.3361 | 0.0004 | 0.3385 | 0.0009 |
| G4 | 0.3641 | 0.0008 | 0.4183 | 0.0009 | 0.3742 | 0.0004 | 0.3668 | 0.0006 | 0.3631 | 0.0014 | 0.3627 | 0.0013 | 0.3320 | 0.0012 | 0.3350 | 0.0020 | 0.3330 | 0.0003 | 0.3368 | 0.0009 |
| G5 | 0.3643 | 0.0010 | 0.4193 | 0.0006 | 0.3777 | 0.0003 | 0.3671 | 0.0005 | 0.3655 | 0.0014 | 0.3654 | 0.0009 | 0.3287 | 0.0011 | 0.3328 | 0.0020 | 0.3304 | 0.0004 | 0.3351 | 0.0010 |
| G6 | 0.3643 | 0.0010 | 0.4201 | 0.0004 | 0.3804 | 0.0006 | 0.3672 | 0.0005 | 0.3681 | 0.0007 | 0.3663 | 0.0004 | 0.3259 | 0.0009 | 0.3313 | 0.0020 | 0.3276 | 0.0002 | 0.3336 | 0.0012 |
| G7 | 0.3643 | 0.0009 | 0.4207 | 0.0006 | 0.3831 | 0.0004 | 0.3675 | 0.0003 | 0.3706 | 0.0009 | 0.3664 | 0.0004 | 0.3237 | 0.0008 | 0.3298 | 0.0023 | 0.3253 | 0.0003 | 0.3322 | 0.0012 |
| G8 | 0.3644 | 0.0007 | 0.4212 | 0.0006 | 0.3856 | 0.0004 | 0.3677 | 0.0004 | 0.3726 | 0.0010 | 0.3665 | 0.0005 | 0.3218 | 0.0009 | 0.3285 | 0.0022 | 0.3233 | 0.0004 | 0.3311 | 0.0016 |
| G9 | 0.3647 | 0.0009 | 0.4216 | 0.0005 | 0.3877 | 0.0003 | 0.3679 | 0.0004 | 0.3745 | 0.0013 | 0.3670 | 0.0004 | 0.3199 | 0.0007 | 0.3275 | 0.0029 | 0.3216 | 0.0005 | 0.3299 | 0.0018 |
| G10 | 0.3651 | 0.0009 | 0.4219 | 0.0006 | 0.3893 | 0.0007 | 0.3681 | 0.0003 | 0.3763 | 0.0012 | 0.3671 | 0.0005 | 0.3183 | 0.0006 | 0.3269 | 0.0030 | 0.3200 | 0.0005 | 0.3288 | 0.0017 |
| G11 | 0.3655 | 0.0008 | 0.4221 | 0.0006 | 0.3908 | 0.0008 | 0.3683 | 0.0004 | 0.3776 | 0.0008 | 0.3675 | 0.0004 | 0.3168 | 0.0006 | 0.3262 | 0.0033 | 0.3187 | 0.0005 | 0.3277 | 0.0017 |
| G12 | 0.3659 | 0.0009 | 0.4223 | 0.0004 | 0.3920 | 0.0010 | 0.3682 | 0.0003 | 0.3793 | 0.0012 | 0.3677 | 0.0004 | 0.3156 | 0.0005 | 0.3255 | 0.0036 | 0.3171 | 0.0006 | 0.3272 | 0.0017 |
| G13 | 0.3664 | 0.0009 | 0.4225 | 0.0004 | 0.3930 | 0.0010 | 0.3684 | 0.0003 | 0.3807 | 0.0011 | 0.3678 | 0.0003 | 0.3143 | 0.0004 | 0.3249 | 0.0034 | 0.3159 | 0.0009 | 0.3268 | 0.0019 |
| G14 | 0.3667 | 0.0010 | 0.4226 | 0.0003 | 0.3940 | 0.0011 | 0.3686 | 0.0005 | 0.3819 | 0.0010 | 0.3682 | 0.0004 | 0.3131 | 0.0002 | 0.3247 | 0.0032 | 0.3148 | 0.0010 | 0.3264 | 0.0019 |
| G15 | 0.3669 | 0.0010 | 0.4228 | 0.0004 | 0.3952 | 0.0013 | 0.3687 | 0.0005 | 0.3832 | 0.0005 | 0.3683 | 0.0003 | 0.3120 | 0.0003 | 0.3249 | 0.0035 | 0.3138 | 0.0010 | 0.3258 | 0.0023 |
| G16 | 0.3673 | 0.0009 | 0.4228 | 0.0004 | 0.3962 | 0.0012 | 0.3688 | 0.0005 | 0.3846 | 0.0004 | 0.3686 | 0.0002 | 0.3107 | 0.0002 | 0.3247 | 0.0038 | 0.3126 | 0.0010 | 0.3256 | 0.0028 |
| G17 | 0.3676 | 0.0010 | 0.4230 | 0.0004 | 0.3970 | 0.0011 | 0.3689 | 0.0005 | 0.3857 | 0.0005 | 0.3688 | 0.0003 | 0.3096 | 0.0002 | 0.3249 | 0.0041 | 0.3116 | 0.0012 | 0.3254 | 0.0030 |
| G18 | 0.3679 | 0.0012 | 0.4231 | 0.0004 | 0.3979 | 0.0012 | 0.3691 | 0.0004 | 0.3867 | 0.0007 | 0.3690 | 0.0003 | 0.3087 | 0.0003 | 0.3251 | 0.0048 | 0.3107 | 0.0014 | 0.3251 | 0.0030 |
| G19 | 0.3685 | 0.0012 | 0.4231 | 0.0003 | 0.3986 | 0.0012 | 0.3692 | 0.0005 | 0.3877 | 0.0007 | 0.3690 | 0.0004 | 0.3077 | 0.0004 | 0.3247 | 0.0050 | 0.3098 | 0.0012 | 0.3251 | 0.0031 |
| G20 | 0.3688 | 0.0012 | 0.4232 | 0.0004 | 0.3992 | 0.0011 | 0.3693 | 0.0005 | 0.3886 | 0.0010 | 0.3693 | 0.0003 | 0.3068 | 0.0005 | 0.3245 | 0.0050 | 0.3091 | 0.0012 | 0.3253 | 0.0030 |
| G21 | 0.3693 | 0.0013 | 0.4232 | 0.0004 | 0.3998 | 0.0012 | 0.3695 | 0.0003 | 0.3895 | 0.0011 | 0.3694 | 0.0004 | 0.3060 | 0.0005 | 0.3241 | 0.0050 | 0.3081 | 0.0013 | 0.3252 | 0.0030 |
| G22 | 0.3697 | 0.0014 | 0.4233 | 0.0004 | 0.4006 | 0.0013 | 0.3696 | 0.0003 | 0.3904 | 0.0010 | 0.3695 | 0.0002 | 0.3053 | 0.0007 | 0.3237 | 0.0052 | 0.3073 | 0.0013 | 0.3257 | 0.0036 |
| G23 | 0.3700 | 0.0014 | 0.4233 | 0.0005 | 0.4012 | 0.0014 | 0.3697 | 0.0003 | 0.3910 | 0.0009 | 0.3695 | 0.0004 | 0.3045 | 0.0007 | 0.3237 | 0.0057 | 0.3066 | 0.0012 | 0.3253 | 0.0035 |
| G24 | 0.3702 | 0.0015 | 0.4233 | 0.0005 | 0.4018 | 0.0015 | 0.3699 | 0.0002 | 0.3919 | 0.0008 | 0.3697 | 0.0004 | 0.3038 | 0.0006 | 0.3234 | 0.0060 | 0.3059 | 0.0012 | 0.3253 | 0.0037 |
| G25 | 0.3706 | 0.0016 | 0.4234 | 0.0004 | 0.4023 | 0.0016 | 0.3700 | 0.0002 | 0.3929 | 0.0011 | 0.3698 | 0.0003 | 0.3029 | 0.0007 | 0.3233 | 0.0062 | 0.3051 | 0.0012 | 0.3250 | 0.0039 |

**Table5.S7** Optimum contribution of Holstein-Friesian cattle to a hypothetical multi-breed population to achieve maximum genetic diversity

| | REF | | MNC | | maxNative.A | | maxNative.B | | maxNative.C | | maxNative.D | | minfCORE.A | | minfCORE.B | | minfCORE.C | | minfCORE.D | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| G0 | 0.0811 | 0.0000 | 0.0939 | 0.0000 | 0.0939 | 0.0000 | 0.0939 | 0.0000 | 0.0939 | 0.0000 | 0.0939 | 0.0000 | 0.0939 | 0.0000 | 0.0939 | 0.0000 | 0.0939 | 0.0000 | 0.0939 | 0.0000 |
| G1 | 0.0821 | 0.0006 | 0.1748 | 0.0002 | 0.1226 | 0.0006 | 0.1224 | 0.0007 | 0.1096 | 0.0006 | 0.1094 | 0.0007 | 0.0907 | 0.0004 | 0.0905 | 0.0010 | 0.0901 | 0.0005 | 0.0899 | 0.0010 |
| G2 | 0.0830 | 0.0014 | 0.1832 | 0.0010 | 0.1309 | 0.0010 | 0.1308 | 0.0012 | 0.1160 | 0.0007 | 0.1157 | 0.0019 | 0.0902 | 0.0010 | 0.0932 | 0.0008 | 0.0894 | 0.0011 | 0.0931 | 0.0017 |
| G3 | 0.0840 | 0.0013 | 0.1858 | 0.0006 | 0.1374 | 0.0006 | 0.1343 | 0.0014 | 0.1214 | 0.0006 | 0.1211 | 0.0020 | 0.0910 | 0.0015 | 0.0947 | 0.0011 | 0.0901 | 0.0009 | 0.0940 | 0.0009 |
| G4 | 0.0848 | 0.0014 | 0.1878 | 0.0008 | 0.1418 | 0.0010 | 0.1364 | 0.0028 | 0.1249 | 0.0012 | 0.1247 | 0.0026 | 0.0917 | 0.0018 | 0.0961 | 0.0022 | 0.0908 | 0.0018 | 0.0957 | 0.0012 |
| G5 | 0.0858 | 0.0016 | 0.1892 | 0.0011 | 0.1456 | 0.0013 | 0.1383 | 0.0021 | 0.1282 | 0.0008 | 0.1287 | 0.0018 | 0.0922 | 0.0019 | 0.0977 | 0.0026 | 0.0911 | 0.0020 | 0.0969 | 0.0010 |
| G6 | 0.0870 | 0.0017 | 0.1903 | 0.0012 | 0.1482 | 0.0014 | 0.1399 | 0.0020 | 0.1317 | 0.0015 | 0.1303 | 0.0022 | 0.0927 | 0.0014 | 0.0992 | 0.0020 | 0.0916 | 0.0021 | 0.0978 | 0.0011 |
| G7 | 0.0881 | 0.0014 | 0.1910 | 0.0011 | 0.1508 | 0.0018 | 0.1410 | 0.0017 | 0.1344 | 0.0022 | 0.1320 | 0.0021 | 0.0938 | 0.0018 | 0.1005 | 0.0021 | 0.0922 | 0.0016 | 0.0986 | 0.0016 |
| G8 | 0.0892 | 0.0011 | 0.1911 | 0.0009 | 0.1535 | 0.0021 | 0.1421 | 0.0020 | 0.1368 | 0.0021 | 0.1331 | 0.0022 | 0.0946 | 0.0015 | 0.1017 | 0.0021 | 0.0926 | 0.0015 | 0.1000 | 0.0017 |
| G9 | 0.0904 | 0.0013 | 0.1917 | 0.0008 | 0.1555 | 0.0021 | 0.1433 | 0.0021 | 0.1391 | 0.0024 | 0.1343 | 0.0025 | 0.0952 | 0.0016 | 0.1030 | 0.0023 | 0.0936 | 0.0013 | 0.1016 | 0.0020 |
| G10 | 0.0916 | 0.0018 | 0.1917 | 0.0009 | 0.1580 | 0.0019 | 0.1448 | 0.0025 | 0.1414 | 0.0019 | 0.1362 | 0.0030 | 0.0958 | 0.0020 | 0.1047 | 0.0022 | 0.0948 | 0.0011 | 0.1022 | 0.0023 |
| G11 | 0.0931 | 0.0018 | 0.1919 | 0.0009 | 0.1595 | 0.0017 | 0.1455 | 0.0031 | 0.1437 | 0.0024 | 0.1371 | 0.0034 | 0.0969 | 0.0018 | 0.1064 | 0.0030 | 0.0956 | 0.0011 | 0.1033 | 0.0018 |
| G12 | 0.0946 | 0.0021 | 0.1920 | 0.0009 | 0.1612 | 0.0019 | 0.1463 | 0.0033 | 0.1463 | 0.0022 | 0.1383 | 0.0036 | 0.0975 | 0.0018 | 0.1072 | 0.0026 | 0.0961 | 0.0009 | 0.1047 | 0.0017 |
| G13 | 0.0960 | 0.0019 | 0.1923 | 0.0010 | 0.1627 | 0.0017 | 0.1471 | 0.0031 | 0.1477 | 0.0017 | 0.1394 | 0.0035 | 0.0980 | 0.0017 | 0.1085 | 0.0025 | 0.0968 | 0.0011 | 0.1064 | 0.0013 |
| G14 | 0.0971 | 0.0023 | 0.1923 | 0.0010 | 0.1638 | 0.0015 | 0.1477 | 0.0032 | 0.1495 | 0.0015 | 0.1406 | 0.0036 | 0.0987 | 0.0019 | 0.1099 | 0.0024 | 0.0974 | 0.0012 | 0.1075 | 0.0013 |
| G15 | 0.0981 | 0.0027 | 0.1924 | 0.0011 | 0.1648 | 0.0014 | 0.1482 | 0.0033 | 0.1508 | 0.0016 | 0.1420 | 0.0032 | 0.0993 | 0.0012 | 0.1116 | 0.0019 | 0.0977 | 0.0015 | 0.1087 | 0.0021 |
| G16 | 0.0994 | 0.0027 | 0.1924 | 0.0010 | 0.1658 | 0.0011 | 0.1489 | 0.0031 | 0.1519 | 0.0016 | 0.1435 | 0.0036 | 0.0996 | 0.0013 | 0.1129 | 0.0017 | 0.0982 | 0.0017 | 0.1101 | 0.0027 |
| G17 | 0.1006 | 0.0028 | 0.1925 | 0.0010 | 0.1666 | 0.0009 | 0.1496 | 0.0027 | 0.1531 | 0.0016 | 0.1451 | 0.0034 | 0.1003 | 0.0019 | 0.1144 | 0.0018 | 0.0990 | 0.0014 | 0.1116 | 0.0030 |
| G18 | 0.1019 | 0.0031 | 0.1925 | 0.0010 | 0.1675 | 0.0010 | 0.1504 | 0.0025 | 0.1544 | 0.0017 | 0.1465 | 0.0029 | 0.1008 | 0.0021 | 0.1154 | 0.0024 | 0.0998 | 0.0011 | 0.1126 | 0.0026 |
| G19 | 0.1032 | 0.0031 | 0.1926 | 0.0010 | 0.1680 | 0.0007 | 0.1511 | 0.0024 | 0.1558 | 0.0016 | 0.1476 | 0.0028 | 0.1014 | 0.0017 | 0.1162 | 0.0024 | 0.1005 | 0.0013 | 0.1133 | 0.0027 |
| G20 | 0.1043 | 0.0034 | 0.1927 | 0.0010 | 0.1690 | 0.0009 | 0.1517 | 0.0026 | 0.1568 | 0.0015 | 0.1488 | 0.0027 | 0.1020 | 0.0019 | 0.1169 | 0.0021 | 0.1010 | 0.0015 | 0.1146 | 0.0029 |
| G21 | 0.1056 | 0.0035 | 0.1926 | 0.0009 | 0.1702 | 0.0009 | 0.1524 | 0.0021 | 0.1578 | 0.0017 | 0.1499 | 0.0024 | 0.1022 | 0.0018 | 0.1177 | 0.0020 | 0.1016 | 0.0014 | 0.1157 | 0.0025 |
| G22 | 0.1065 | 0.0035 | 0.1927 | 0.0009 | 0.1708 | 0.0007 | 0.1530 | 0.0020 | 0.1586 | 0.0020 | 0.1506 | 0.0020 | 0.1028 | 0.0017 | 0.1186 | 0.0020 | 0.1019 | 0.0018 | 0.1169 | 0.0025 |
| G23 | 0.1073 | 0.0035 | 0.1927 | 0.0009 | 0.1714 | 0.0005 | 0.1537 | 0.0018 | 0.1594 | 0.0020 | 0.1509 | 0.0019 | 0.1036 | 0.0017 | 0.1198 | 0.0024 | 0.1031 | 0.0020 | 0.1172 | 0.0025 |
| G24 | 0.1083 | 0.0039 | 0.1927 | 0.0009 | 0.1719 | 0.0003 | 0.1542 | 0.0018 | 0.1601 | 0.0020 | 0.1518 | 0.0019 | 0.1039 | 0.0016 | 0.1206 | 0.0026 | 0.1039 | 0.0021 | 0.1180 | 0.0026 |
| G25 | 0.1090 | 0.0037 | 0.1927 | 0.0009 | 0.1725 | 0.0005 | 0.1544 | 0.0019 | 0.1613 | 0.0020 | 0.1520 | 0.0016 | 0.1045 | 0.0019 | 0.1211 | 0.0028 | 0.1047 | 0.0022 | 0.1184 | 0.0030 |

**Table5.S8** Optimum contribution of Red Holstein cattle to a hypothetical multi-breed population to achieve maximum genetic diversity

| | REF | | MNC | | maxNative.A | | maxNative.B | | maxNative.C | | maxNative.D | | minfCORE.A | | minfCORE.B | | minfCORE.C | | minfCORE.D | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| G0 | 0.0200 | 0.0000 | 0.0048 | 0.0000 | 0.0048 | 0.0000 | 0.0048 | 0.0000 | 0.0048 | 0.0000 | 0.0048 | 0.0000 | 0.0048 | 0.0000 | 0.0048 | 0.0000 | 0.0048 | 0.0000 | 0.0048 | 0.0000 |
| G1 | 0.0200 | 0.0000 | 0.0568 | 0.0002 | 0.0229 | 0.0003 | 0.0229 | 0.0005 | 0.0140 | 0.0006 | 0.0140 | 0.0006 | 0.0015 | 0.0004 | 0.0017 | 0.0005 | 0.0012 | 0.0002 | 0.0017 | 0.0006 |
| G2 | 0.0200 | 0.0000 | 0.0632 | 0.0014 | 0.0293 | 0.0007 | 0.0293 | 0.0006 | 0.0185 | 0.0007 | 0.0185 | 0.0011 | 0.0039 | 0.0012 | 0.0056 | 0.0012 | 0.0022 | 0.0011 | 0.0058 | 0.0010 |
| G3 | 0.0200 | 0.0000 | 0.0661 | 0.0011 | 0.0332 | 0.0013 | 0.0322 | 0.0018 | 0.0210 | 0.0010 | 0.0212 | 0.0013 | 0.0043 | 0.0009 | 0.0078 | 0.0008 | 0.0030 | 0.0016 | 0.0079 | 0.0013 |
| G4 | 0.0200 | 0.0000 | 0.0676 | 0.0010 | 0.0363 | 0.0014 | 0.0335 | 0.0023 | 0.0241 | 0.0012 | 0.0243 | 0.0014 | 0.0054 | 0.0012 | 0.0085 | 0.0009 | 0.0044 | 0.0025 | 0.0095 | 0.0012 |
| G5 | 0.0200 | 0.0000 | 0.0681 | 0.0009 | 0.0395 | 0.0016 | 0.0345 | 0.0020 | 0.0269 | 0.0021 | 0.0265 | 0.0013 | 0.0061 | 0.0016 | 0.0098 | 0.0007 | 0.0058 | 0.0020 | 0.0110 | 0.0014 |
| G6 | 0.0200 | 0.0000 | 0.0684 | 0.0010 | 0.0422 | 0.0014 | 0.0357 | 0.0017 | 0.0290 | 0.0021 | 0.0286 | 0.0011 | 0.0076 | 0.0018 | 0.0111 | 0.0010 | 0.0066 | 0.0017 | 0.0125 | 0.0014 |
| G7 | 0.0200 | 0.0000 | 0.0688 | 0.0007 | 0.0448 | 0.0015 | 0.0373 | 0.0017 | 0.0312 | 0.0019 | 0.0304 | 0.0011 | 0.0086 | 0.0019 | 0.0125 | 0.0011 | 0.0077 | 0.0018 | 0.0140 | 0.0013 |
| G8 | 0.0200 | 0.0000 | 0.0694 | 0.0005 | 0.0466 | 0.0016 | 0.0387 | 0.0017 | 0.0328 | 0.0025 | 0.0324 | 0.0017 | 0.0095 | 0.0017 | 0.0135 | 0.0014 | 0.0090 | 0.0018 | 0.0152 | 0.0015 |
| G9 | 0.0200 | 0.0000 | 0.0697 | 0.0007 | 0.0481 | 0.0017 | 0.0399 | 0.0016 | 0.0347 | 0.0023 | 0.0343 | 0.0012 | 0.0102 | 0.0020 | 0.0146 | 0.0015 | 0.0099 | 0.0019 | 0.0162 | 0.0011 |
| G10 | 0.0200 | 0.0000 | 0.0702 | 0.0008 | 0.0491 | 0.0015 | 0.0405 | 0.0020 | 0.0365 | 0.0015 | 0.0355 | 0.0019 | 0.0109 | 0.0023 | 0.0155 | 0.0017 | 0.0106 | 0.0017 | 0.0173 | 0.0013 |
| G11 | 0.0200 | 0.0000 | 0.0704 | 0.0008 | 0.0505 | 0.0013 | 0.0417 | 0.0018 | 0.0375 | 0.0012 | 0.0369 | 0.0020 | 0.0113 | 0.0025 | 0.0163 | 0.0016 | 0.0111 | 0.0015 | 0.0179 | 0.0014 |
| G12 | 0.0200 | 0.0000 | 0.0706 | 0.0010 | 0.0516 | 0.0013 | 0.0428 | 0.0020 | 0.0385 | 0.0014 | 0.0382 | 0.0022 | 0.0122 | 0.0020 | 0.0176 | 0.0019 | 0.0120 | 0.0017 | 0.0186 | 0.0011 |
| G13 | 0.0200 | 0.0000 | 0.0707 | 0.0010 | 0.0524 | 0.0010 | 0.0435 | 0.0018 | 0.0401 | 0.0009 | 0.0392 | 0.0023 | 0.0130 | 0.0018 | 0.0179 | 0.0020 | 0.0127 | 0.0017 | 0.0191 | 0.0011 |
| G14 | 0.0200 | 0.0000 | 0.0708 | 0.0012 | 0.0534 | 0.0010 | 0.0443 | 0.0016 | 0.0408 | 0.0006 | 0.0404 | 0.0023 | 0.0137 | 0.0017 | 0.0186 | 0.0020 | 0.0135 | 0.0020 | 0.0200 | 0.0011 |
| G15 | 0.0200 | 0.0000 | 0.0710 | 0.0013 | 0.0545 | 0.0009 | 0.0452 | 0.0018 | 0.0419 | 0.0008 | 0.0409 | 0.0024 | 0.0144 | 0.0017 | 0.0194 | 0.0029 | 0.0145 | 0.0018 | 0.0209 | 0.0014 |
| G16 | 0.0200 | 0.0000 | 0.0710 | 0.0013 | 0.0555 | 0.0010 | 0.0459 | 0.0016 | 0.0436 | 0.0007 | 0.0413 | 0.0027 | 0.0153 | 0.0017 | 0.0199 | 0.0032 | 0.0151 | 0.0017 | 0.0213 | 0.0012 |
| G17 | 0.0200 | 0.0000 | 0.0712 | 0.0013 | 0.0563 | 0.0009 | 0.0465 | 0.0015 | 0.0445 | 0.0010 | 0.0419 | 0.0029 | 0.0157 | 0.0016 | 0.0207 | 0.0035 | 0.0153 | 0.0019 | 0.0217 | 0.0012 |
| G18 | 0.0200 | 0.0000 | 0.0712 | 0.0013 | 0.0569 | 0.0006 | 0.0467 | 0.0016 | 0.0452 | 0.0008 | 0.0421 | 0.0024 | 0.0164 | 0.0018 | 0.0217 | 0.0037 | 0.0156 | 0.0017 | 0.0225 | 0.0010 |
| G19 | 0.0200 | 0.0000 | 0.0712 | 0.0013 | 0.0577 | 0.0003 | 0.0471 | 0.0015 | 0.0461 | 0.0008 | 0.0426 | 0.0025 | 0.0167 | 0.0015 | 0.0226 | 0.0041 | 0.0160 | 0.0015 | 0.0236 | 0.0010 |
| G20 | 0.0200 | 0.0000 | 0.0713 | 0.0013 | 0.0582 | 0.0005 | 0.0474 | 0.0018 | 0.0470 | 0.0008 | 0.0430 | 0.0023 | 0.0172 | 0.0021 | 0.0234 | 0.0043 | 0.0166 | 0.0015 | 0.0241 | 0.0011 |
| G21 | 0.0200 | 0.0001 | 0.0713 | 0.0013 | 0.0583 | 0.0003 | 0.0479 | 0.0015 | 0.0479 | 0.0009 | 0.0435 | 0.0023 | 0.0180 | 0.0016 | 0.0240 | 0.0043 | 0.0171 | 0.0014 | 0.0245 | 0.0012 |
| G22 | 0.0201 | 0.0002 | 0.0714 | 0.0012 | 0.0591 | 0.0004 | 0.0484 | 0.0014 | 0.0488 | 0.0012 | 0.0440 | 0.0022 | 0.0185 | 0.0017 | 0.0244 | 0.0046 | 0.0177 | 0.0014 | 0.0253 | 0.0011 |
| G23 | 0.0203 | 0.0007 | 0.0714 | 0.0012 | 0.0595 | 0.0004 | 0.0488 | 0.0012 | 0.0497 | 0.0013 | 0.0449 | 0.0019 | 0.0189 | 0.0020 | 0.0250 | 0.0046 | 0.0176 | 0.0015 | 0.0261 | 0.0010 |
| G24 | 0.0203 | 0.0006 | 0.0715 | 0.0011 | 0.0599 | 0.0005 | 0.0493 | 0.0010 | 0.0507 | 0.0013 | 0.0454 | 0.0016 | 0.0197 | 0.0021 | 0.0255 | 0.0046 | 0.0181 | 0.0011 | 0.0269 | 0.0013 |
| G25 | 0.0204 | 0.0008 | 0.0715 | 0.0011 | 0.0603 | 0.0006 | 0.0498 | 0.0013 | 0.0510 | 0.0015 | 0.0460 | 0.0014 | 0.0200 | 0.0023 | 0.0263 | 0.0045 | 0.0184 | 0.0010 | 0.0277 | 0.0016 |

**Table5.S9** Optimum contribution of Red Holstein cattle to a hypothetical multi-breed population to achieve maximum genetic diversity

| | REF | | MNC | | maxNative.A | | maxNative.B | | maxNative.C | | maxNative.D | | minfCORE.A | | minfCORE.B | | minfCORE.C | | minfCORE.D | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| G0 | 0.2074 | 0.0000 | 0.2071 | 0.0000 | 0.2071 | 0.0000 | 0.2071 | 0.0000 | 0.2071 | 0.0000 | 0.2071 | 0.0000 | 0.2071 | 0.0000 | 0.2071 | 0.0000 | 0.2071 | 0.0000 | 0.2071 | 0.0000 |
| G1 | 0.2086 | 0.0007 | 0.2671 | 0.0001 | 0.2089 | 0.0004 | 0.2088 | 0.0004 | 0.1966 | 0.0006 | 0.1965 | 0.0006 | 0.1888 | 0.0005 | 0.1886 | 0.0007 | 0.1891 | 0.0005 | 0.1891 | 0.0003 |
| G2 | 0.2092 | 0.0008 | 0.2790 | 0.0016 | 0.2153 | 0.0004 | 0.2151 | 0.0006 | 0.1997 | 0.0007 | 0.1997 | 0.0008 | 0.1857 | 0.0007 | 0.1870 | 0.0011 | 0.1864 | 0.0005 | 0.1875 | 0.0015 |
| G3 | 0.2096 | 0.0007 | 0.2843 | 0.0017 | 0.2225 | 0.0009 | 0.2207 | 0.0014 | 0.2049 | 0.0008 | 0.2055 | 0.0005 | 0.1841 | 0.0007 | 0.1864 | 0.0009 | 0.1845 | 0.0009 | 0.1873 | 0.0012 |
| G4 | 0.2100 | 0.0007 | 0.2867 | 0.0014 | 0.2290 | 0.0012 | 0.2238 | 0.0013 | 0.2096 | 0.0014 | 0.2098 | 0.0005 | 0.1833 | 0.0008 | 0.1870 | 0.0010 | 0.1835 | 0.0009 | 0.1877 | 0.0003 |
| G5 | 0.2107 | 0.0007 | 0.2882 | 0.0014 | 0.2343 | 0.0011 | 0.2258 | 0.0013 | 0.2136 | 0.0011 | 0.2142 | 0.0011 | 0.1829 | 0.0009 | 0.1872 | 0.0008 | 0.1832 | 0.0004 | 0.1882 | 0.0006 |
| G6 | 0.2111 | 0.0005 | 0.2891 | 0.0013 | 0.2388 | 0.0007 | 0.2274 | 0.0013 | 0.2174 | 0.0013 | 0.2170 | 0.0016 | 0.1824 | 0.0007 | 0.1875 | 0.0011 | 0.1827 | 0.0005 | 0.1887 | 0.0008 |
| G7 | 0.2117 | 0.0007 | 0.2901 | 0.0009 | 0.2426 | 0.0006 | 0.2292 | 0.0012 | 0.2215 | 0.0021 | 0.2187 | 0.0015 | 0.1823 | 0.0005 | 0.1882 | 0.0019 | 0.1827 | 0.0007 | 0.1895 | 0.0010 |
| G8 | 0.2121 | 0.0007 | 0.2907 | 0.0008 | 0.2458 | 0.0009 | 0.2307 | 0.0012 | 0.2248 | 0.0020 | 0.2208 | 0.0017 | 0.1825 | 0.0006 | 0.1888 | 0.0018 | 0.1828 | 0.0008 | 0.1901 | 0.0013 |
| G9 | 0.2129 | 0.0010 | 0.2912 | 0.0007 | 0.2490 | 0.0009 | 0.2320 | 0.0011 | 0.2277 | 0.0021 | 0.2224 | 0.0017 | 0.1828 | 0.0004 | 0.1897 | 0.0025 | 0.1829 | 0.0007 | 0.1908 | 0.0016 |
| G10 | 0.2135 | 0.0010 | 0.2916 | 0.0007 | 0.2513 | 0.0012 | 0.2331 | 0.0011 | 0.2303 | 0.0024 | 0.2242 | 0.0014 | 0.1832 | 0.0006 | 0.1908 | 0.0025 | 0.1829 | 0.0007 | 0.1916 | 0.0018 |
| G11 | 0.2143 | 0.0010 | 0.2919 | 0.0007 | 0.2533 | 0.0013 | 0.2341 | 0.0005 | 0.2328 | 0.0021 | 0.2261 | 0.0018 | 0.1833 | 0.0008 | 0.1916 | 0.0028 | 0.1832 | 0.0008 | 0.1922 | 0.0016 |
| G12 | 0.2150 | 0.0012 | 0.2921 | 0.0006 | 0.2551 | 0.0014 | 0.2351 | 0.0005 | 0.2354 | 0.0022 | 0.2276 | 0.0018 | 0.1839 | 0.0007 | 0.1926 | 0.0028 | 0.1834 | 0.0008 | 0.1929 | 0.0015 |
| G13 | 0.2157 | 0.0012 | 0.2923 | 0.0005 | 0.2569 | 0.0012 | 0.2359 | 0.0007 | 0.2377 | 0.0021 | 0.2290 | 0.0015 | 0.1840 | 0.0008 | 0.1933 | 0.0028 | 0.1838 | 0.0007 | 0.1936 | 0.0015 |
| G14 | 0.2163 | 0.0011 | 0.2925 | 0.0006 | 0.2585 | 0.0009 | 0.2369 | 0.0007 | 0.2395 | 0.0019 | 0.2302 | 0.0016 | 0.1843 | 0.0009 | 0.1945 | 0.0025 | 0.1841 | 0.0010 | 0.1943 | 0.0016 |
| G15 | 0.2169 | 0.0013 | 0.2926 | 0.0005 | 0.2599 | 0.0011 | 0.2379 | 0.0009 | 0.2419 | 0.0019 | 0.2313 | 0.0014 | 0.1845 | 0.0009 | 0.1956 | 0.0030 | 0.1841 | 0.0010 | 0.1949 | 0.0021 |
| G16 | 0.2175 | 0.0016 | 0.2927 | 0.0005 | 0.2613 | 0.0010 | 0.2387 | 0.0009 | 0.2439 | 0.0019 | 0.2329 | 0.0012 | 0.1845 | 0.0010 | 0.1966 | 0.0033 | 0.1843 | 0.0009 | 0.1957 | 0.0027 |
| G17 | 0.2182 | 0.0017 | 0.2928 | 0.0005 | 0.2625 | 0.0010 | 0.2395 | 0.0008 | 0.2455 | 0.0021 | 0.2340 | 0.0010 | 0.1846 | 0.0008 | 0.1978 | 0.0035 | 0.1845 | 0.0009 | 0.1965 | 0.0029 |
| G18 | 0.2186 | 0.0017 | 0.2929 | 0.0005 | 0.2635 | 0.0008 | 0.2402 | 0.0009 | 0.2470 | 0.0022 | 0.2349 | 0.0011 | 0.1849 | 0.0009 | 0.1988 | 0.0040 | 0.1850 | 0.0011 | 0.1974 | 0.0028 |
| G19 | 0.2195 | 0.0019 | 0.2930 | 0.0005 | 0.2646 | 0.0008 | 0.2407 | 0.0008 | 0.2485 | 0.0021 | 0.2355 | 0.0009 | 0.1852 | 0.0008 | 0.1994 | 0.0039 | 0.1851 | 0.0009 | 0.1983 | 0.0030 |
| G20 | 0.2200 | 0.0016 | 0.2931 | 0.0005 | 0.2655 | 0.0008 | 0.2414 | 0.0008 | 0.2499 | 0.0023 | 0.2366 | 0.0009 | 0.1853 | 0.0005 | 0.2003 | 0.0039 | 0.1853 | 0.0009 | 0.1993 | 0.0029 |
| G21 | 0.2207 | 0.0017 | 0.2931 | 0.0004 | 0.2664 | 0.0010 | 0.2421 | 0.0005 | 0.2513 | 0.0023 | 0.2373 | 0.0007 | 0.1856 | 0.0005 | 0.2009 | 0.0041 | 0.1855 | 0.0010 | 0.2001 | 0.0028 |
| G22 | 0.2212 | 0.0018 | 0.2932 | 0.0004 | 0.2672 | 0.0010 | 0.2426 | 0.0007 | 0.2526 | 0.0021 | 0.2381 | 0.0007 | 0.1858 | 0.0007 | 0.2012 | 0.0042 | 0.1856 | 0.0011 | 0.2011 | 0.0033 |
| G23 | 0.2218 | 0.0020 | 0.2933 | 0.0003 | 0.2682 | 0.0011 | 0.2430 | 0.0005 | 0.2538 | 0.0018 | 0.2389 | 0.0006 | 0.1859 | 0.0008 | 0.2019 | 0.0046 | 0.1858 | 0.0011 | 0.2014 | 0.0031 |
| G24 | 0.2222 | 0.0019 | 0.2934 | 0.0003 | 0.2690 | 0.0011 | 0.2435 | 0.0006 | 0.2548 | 0.0017 | 0.2398 | 0.0006 | 0.1861 | 0.0008 | 0.2024 | 0.0048 | 0.1860 | 0.0011 | 0.2021 | 0.0031 |
| G25 | 0.2227 | 0.0020 | 0.2935 | 0.0004 | 0.2697 | 0.0010 | 0.2441 | 0.0007 | 0.2561 | 0.0018 | 0.2404 | 0.0006 | 0.1862 | 0.0009 | 0.2031 | 0.0050 | 0.1861 | 0.0011 | 0.2027 | 0.0033 |

# CHAPTER 6

# GENERAL DISCUSSION

# 6.1 Introduction

In the breeding schemes of small local breeds, the conflicts of breeding goals must be addressed: the conflict between increasing genetic gain while managing the inbreeding level and the conflict between maintaining genetic diversity while controlling the loss of genetic originality. Traditional optimum contribution selection (traditional OCS) aims at achieving the maximum genetic gain whilst controlling the rate of inbreeding when making decisions on the number of matings to allocate to each selection candidate (Meuwissen 1997; Woolliams *et al.* 2015). To date, OCS has become a well-established method that has been implemented in animal breeding schemes (Doekes *et al.* 2018). It has been proved to be efficient in maintaining genetic diversity since it aims at minimizing or restricting average relatedness of the potential parents in order to minimize the rate of inbreeding. Thus the genetic diversity in the long term is conserved. However, this diversity may be caused by high amount of genetic material brought by the foreign ancestors due to introgression process, which in turn may threaten the conservation of small local populations. Amador *et al.* (2011) confirmed that after several generations without efficient management, even a small introduction of foreign genetic material will rapidly disperse throughout the original population, and this material is difficult to remove. Therefore, the current breeding objectives should not only focus on increasing genetic gain but also maintaining genetic originality and native allele diversity.

The main objective of this thesis was to develop selection strategies that can balance the need for increasing genetic gain, maintaining genetic diversity and genetic originality to enable more efficient long-term management in both purposes of production and conservation in the breeding program of local livestock breeds. In **chapter 2** we evaluated the current inbreeding level based on both pedigree and genomic information of 182 German Angler cattle**.** Due to the gene flow caused by the historical introgression, the current inbreeding level was relatively low based on all three approaches ($\overline{F_{PED}}$=0.013; $\overline{F_{GRM}}$=−0.015; $\overline{F_{ROH>1Mb}}$=0.031). Although traditional OCS may achieve higher genetic gain with the restriction of the defined rate of inbreeding, in this case, inbreeding is not the main problem in the current breeding program. Thus the advantage of OCS may be limited since the level of inbreeding may be lower than the threshold. On the other hand, the inbreeding coefficients estimated based on runs of homozygosity (ROH) has a good reflection of true inbreeding level since it captures both recent and ancient inbreeding (Ferenčaković *et al.* 2013). In **chapter 4** and **chapter 5** instead of using

classic genomic relationship matrix (G-matrix, VanRaden 2008), the kinship among selection candidate based on ROH was used for the optimization.

After the top Red Holstein sires were introduced in the Angler breeding program since the 1980s, an additional breeding goal should be considered is to recover the original genetic background of Angler breed and reduce the genetic material introduced from Red Holstein in the population. The performance of advanced OCS, which considers migrant contribution and kinship at native alleles as additional constraints in OCS in both production and conservation function was carried out and tested in **chapter 3** based on the pedigree information. Both objective functions of production and conservation were balanced via considering a set of appropriate constraints in the optimization process. The advance OCS was proved to be efficient in balancing both functions of production and conservation under the circumstances of an appropriate setting of constraint values. Subsequently, we evaluated the long-term performance of the advanced OCS strategies in achieving a high genetic gain in **chapter 4**. Advanced OCS outperformed than traditional OCS in maintaining genetic originality when the aim was to achieve the highest genetic gain in the breeding scheme of Angler. In **chapter 5** we imitated a conservation program and evaluated the long-term performance of advanced OCS strategies in recovering the native genetic background, where no production performance was taken into consideration. In the perspective of conservation, advanced OCS also has potential to recover the native genetic background and increase the genetic diversity among breeds, which was not possible to be achieved via using traditional OCS.

This general discussion gives a brief introduction of the current algorithm for solving the optimization problems, utilizing genomic selection in small local breeds, OCS and mating strategies in the genomic era and conclusion remarks.

## 6.2 Algorithm for solving the optimization problems

The classic optimization problem for OCS is to maximize the genetic gain of the following generation $EBV_{t+1}$ whilst constrain the average coancestry $C_{t+1}$. The vector $\mathbf{c_t}$ of selection candidates' genetic contributions in generation $t$ is optimized to maximize $EBV_{t+1} = \mathbf{c}_t^T\widehat{\mathbf{EBV}}_t$ given that $\mathbf{c}_t^T\mathbf{A}_t\mathbf{c}_t/2 = C_{t+1}$, where $\mathbf{A}$ is the relationship matrix of the selection candidates in generation $t$. The genetic contribution $c_i$ of animal i is the fraction of genes in the next generation that originate from this individual. Genetic contributions cannot be negative, i.e. $c_i \geq 0$. The total genetic contribution of candidates from each sex must be equal to 0.5 for diploid species, i.e.$\mathbf{Q}^T\mathbf{c}_t = 1/2\,\mathbf{1}$ , where $\mathbf{Q}$ is an incidence matrix indicating the candidate's sex with 0's and 1's.

Originally the constrained optimization problem was solved by an unconstrained maximization of the Lagrangian function, which is implemented in the software GENCONT (Meuwissen 1997, 2002). $H_t$ is maximized for $\mathbf{c_t}$, $\lambda_0$ and $\boldsymbol{\lambda}$:

$$H_t = \mathbf{c}_t'\mathbf{EBV}_t - (\mathbf{c}_t'\mathbf{A}_t\mathbf{c}_t - 2\bar{C}_{t+1})\lambda_0 - (\mathbf{Q}^T\mathbf{c}_t - 1/2\,\mathbf{1})^T\boldsymbol{\lambda}$$

where $\lambda_0$ and $\boldsymbol{\lambda}$ are Lagrangian multipliers. The optimum solutions are $\mathbf{c_t} = \mathbf{A_t^{-1}}(\mathbf{EBV_t} - \mathbf{Q}\boldsymbol{\lambda})/2\lambda_0$. One drawback of Lagrangian multipliers method is that this algorithm does not always generate optimal solutions. Moreover, it requires the inverse of the relationship matrix of the selection candidates, which is a challenge due to computational limitation when the breeding scheme involves a large number of selection candidates. Dagnachew and Meuwissen (2016) developed a novel iterative algorithm that avoids the inversion of the relationship matrices, which is implemented in the software GENCONT2 (Dagnachew and Meuwissen 2014). It achieved a reduction in computing time of 90-93% compared to the original algorithm and was able to handle datasets consist of a large number of selection candidates. In addition, several algorithms are also frequently used for solving the optimization problems such as: evolutionary algorithms implemented in the software EVA (Berg *et al.* 2006), semidefinite programming (Pong-Wong and Woolliams 2007), branch-and-bound algorithms (Mullin and Belotti 2016) implemented in the software OPSEL (Mullin 2017) etc.

The algorithms mentioned above are most frequently used in solving the optimization problems in the animal and plant breeding area. However, the common drawback of these algorithms is that these algorithms are restricted to only solve the problem between EBV and the coancestry. There is no possibility to include additional constraints in the optimization process, which is unsatisfactory to handle the realistic situation of different population. In **chapter 3**, **chapter 4** and **chapter 5**, we solved the optimization problems using algorithm cccp (Pfaff 2014) called from R package *optiSel* (Wellmann 2018), which is capable of including multiple constraints in the optimization problems. In this project, our main objective was to balance the functions of both production and conservation in the breeding program. Thus, the migrant contribution was taken into consideration as both additional constraint and objective function. In practical, except for migrant contribution, it is also possible to take other realistic parameters into account according to practical needs while designing the breeding programs. Moreover, compared to the other packages, *optiSel* has the advantage of a faster computing speed and the ability to handle a large number of selection candidates. Thus the package *optiSel* is desirable for dealing with breeding programs with multiple realistic breeding goals.

## 6.3 Genomic selection in small local breeds

Since the last decade, the technological improvement has allowed for animals to have genome-wide dense marker genotype data. For commercial cattle breeds, genomic evaluations are routinely performed using the genotype of ~54,000 single nucleotide polymorphisms (SNPs) around the world (e.g., Holstein-Friesian – US, UK, Canada, the Netherlands, Germany, Australia, New Zealand; Fleckvieh – Germany, Austria, Switzerland; Brown Swiss – US, Canada, Austria, Germany, Switzerland etc.). However, for the small local breeds, the procedure may not be routinely performed. Even if genomic selection methods may produce uncertain results in small breeds, still there are several reasons showing why the routinely genotyping animals of the small population can be beneficial (Mészáros *et al.* 2015). In the genomic selection era, using SNP information is prevailing for constructing kinship matrix and breeding value estimation. The genomic information matrix can provide a better estimation of relatedness within the population and uncover the family structures, even if there is a lack of complete pedigree information (Calus *et al.* 2011). In addition, for the traditional progeny testing scheme, there is limited opportunity for the small breeding population to achieve a sufficiently large number of daughters to generate highly reliable EBVs. The problem becomes more serious when the heritability of the trait is low. The accuracy of predicting EBV via using

genomic information could be significantly improved compared to the progeny tested EBVs. It has been reported that the genomic breeding scheme was both genetically and economically superior to the conventional breeding scheme (Thomasen *et al.* 2014a).

However, there are still several problems of applying genomic selection in the small local cattle population. The accuracy of the breeding value estimation based on genomic information depends heavily on the size of the reference population (Goddard and Hayes 2009; Liu *et al.* 2011). With the collaborations on both national and international level, it is possible to generate large reference population for the commercial breeds such as the Holstein breed (Lund *et al.* 2011). However for the small local breed, due to the limited size of the reference population, the accuracy of genomic prediction may not be sufficiently high. Thus it poses a challenge in the future genetic improvement relatively compared to the breed with a larger reference population.

Several approaches could be utilized to solve this problem (Lund *et al.* 2016). For dairy cattle, several studies pointed out that it is worthwhile to genotype not only bulls but also cows (Calus *et al.* 2013; Thomasen *et al.* 2014b; Gao *et al.* 2015). Given the small population size of local breed, the cost for genotyping all animals in the population is smaller compared to the commercial breeds. The second strategy was to combine the reference data from different populations of the same breed. The extent of the improvement of the prediction accuracy may be estimated using the formulas that describe the relationship between the number of individuals in the reference populations and accuracies (Goddard and Hayes 2009). Due to the genetic material exchange between the populations, there may be a significantly high relationship among the individuals from the different population. The accuracy of genomic prediction of the given population will thus be improved via including those populations in the reference population. This is particularly clear for the Nordic red cattle breeds: Danish Red (DR), Swedish Red (SR), Finish Ayrshire (FA) and Norwegian Red (NR). Table 6.1 presents the results from genomic predictions by combining different red cattle populations from Lund *et al.* (2014). Majority of the prediction accuracy of both production and health traits were improved when going from single breed reference population to a joint reference population.

It is still a debate whether including distant-related breeds could improve the accuracy of genomic prediction. Several studies have reported that when using 54K data and GBLUP methods, combining Holstein-Frisian (HF) cattle and Jersey cattle in a joint reference population have little improvements in the prediction accuracy of either HF or Jersey animals (Hayes *et al.* 2009; Olson *et al.* 2012; Erbe *et al.* 2014). Recently, Raymond *et al.* (2018) showed that using GBLUP model results in a very low accuracy of across-breed genomic

prediction in a joint population consisting New Zealand Jersey, New Zealand Holstein and Dutch Holstein animals. The low accuracy of across breeds genomic prediction may be due to the inconsistency of linkage disequilibrium between markers and causal mutations across breeds (Meuwissen *et al.* 2001). Alternatively, when using denser SNP panels, functional subsets of markers or Bayesian methods, the prediction accuracy of EBV of Jersey animals are higher when HF animals are included in the reference population (Hayes *et al.* 2009; Erbe *et al.* 2014).

Currently, bovine whole genome sequence (WGS) data are becoming available. Because of the 1000 bulls genome project (Daetwyler *et al.* 2014), it is possible for the genotyped animals to be imputed to whole genome sequence data via using the sequenced bulls as the reference panel. The application of WGS data has the potential to increase the accuracy of genomic prediction because it is expected to contain all the causal mutations. According to the simulation study of Iheshiulor *et al.* (2016), the reliability of genomic prediction can be increased via using sequence information. However currently, there are still limitations in this approach. One is that the reliability of the imputation is not high enough to capture all rare variances (Daetwyler *et al.* 2014). Another limitation is that it is hard for Bayesian variable selection models to handle large amount (>20M) of SNPs in the prediction process, especially in large populations. To conquer above limitation, sequence many more animals may be required to capture rare causative variance.

In **chapter 4**, we evaluated the relationship between Angler cattle and other breeds based on genomic information (Holstein-Frisian cattle, Red Holstein cattle, Fleckvieh cattle and Norwegian Red cattle) via constructing the kinship matrix and principal component analysis (PCA). Both results showed that Angler cattle has a close relationship with Red Holstein and Holstein-Frisian cattle, a moderate close relationship with Norwegian Red and a distance relationship with Fleckvieh. Currently, with the collaboration between Viking Genetics and VIT (Verden), it is possible to execute the genomic evaluation for the Angler breed (GGI - German Genetics International GmbH). The genomic information of Angler bulls are included in the gene pool along with the genomic information of other red cattle from Nordic countries (Denmark, Sweden, Norway and Finland) to form the reference population. Subsequently, the genomic estimated breeding values (GEBV) became available for Angler cattle with a higher reliability (Stålhammar 2018). It is promising to include animals from other red cattle breeds in the joint population as it is convinced in the recent research project "ReDiverse" (Hinrichs *et al.* 2017). Furthermore, due to the genetic material exchange between Holstein and Angler cattle, including Holstein cattle in the reference population may also be a promising approach to improve the genomic prediction accuracy.

**Table 6.1** Increase in accuracy/reliability when using joint dairy reference compared to a single reference population for milk-, protein and fat yield, fertility and Somatic Cell Score (SCS). All studies are performed using 54k genotype data. Ref1 is the breed and country of origin for the single reference population, and Ref2 is the breeds and countries of origin for the joint reference. Reference sizes are given as number of bulls. R or $R^2$ in column five states whether the original paper uses the correlation or squared correlation to measure the validation accuracy. Breed codes: HF=Holstein-Friesian, DR=Danish Red, SR=Swedish Red, FA=Finnish Ayrshire, NR=Norwegian Red, VR=Danish/Swedish/Finnish Red. Country codes: NO: Nordic, Trait codes: NRR= Non-Return Rate, UHI=Udder Health Index.[1]

| Ref1 | Ref 2 | Ref1 size | Ref2 size | | Milk | Protein | Fat | Fertility | SCS | Method | Citation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DR | VR | 929 | 3735 | $R^2$ | 2 | 4 | 1 | $-3^{NRR}$ | $2^{UHI}$ | Bayesian | Brøndum *et al.* (2011) |
| SR | VR | 1551 | 3735 | $R^2$ | 9 | 18 | 7 | $9^{NRR}$ | $6^{UHI}$ | Bayesian | Brøndum *et al.* (2011) |
| FA | VR | 1562 | 3735 | $R^2$ | 12 | 13 | 6 | $5^{NRR}$ | $10^{UHI}$ | Bayesian | Brøndum *et al.* (2011) |
| VR | VR+NR | 3367 | 5717 | R | 1 | 1 | 2 | $0^{NRR}$ | $2^{UHI}$ | GBLUP | Zhou *et al.* (2014a) |
| NR | VR+NR | 2076 | 5433 | R | 5 | 8 | 5 | $2^{NRR}$ | | GBLUP | Zhou *et al.* (2014a) |
| VR | VR+HF(NO) | 3437 | 6552 | R | 1.4 | 1.1 | 1.0 | $0.4^{NRR}$ | 0.4 | GBLUP | Zhou *et al.* (2014b) |
| DR | VR+HF(NO) | 3437 | 6552 | R | 5 | 3 | 2 | $2^{NRR}$ | 1 | GBLUP | Zhou *et al.* (2014b) |
| SR | VR+HF(NO) | 3437 | 6552 | R | 2 | 2 | 2 | $0^{NRR}$ | 0 | GBLUP | Zhou *et al.* (2014b) |
| FA | VR+HF(NO) | 3437 | 6552 | R | 1 | 0 | 0 | $0^{NRR}$ | 0 | GBLUP | Zhou *et al.* (2014b) |
| HF(NO) | VR+HF(NO) | 3115 | 6552 | R | 0.6 | 0 | 0.4 | $-0.4^{NRR}$ | 0.4 | GBLUP | Zhou *et al.* (2014b) |

[1] Part of Table 1 from Lund *et al.* (2014)

# 6.4 OCS and mating strategies in the genomic era

## 6.4.1 OCS in the genomic era

In the genomics era, the OCS strategy has a further extension due to the availability of marker data for the selection candidates: genomic OCS (GOCS, Sonesson *et al.* 2012). The core of OCS was to calculate the average coancestry $\mathbf{c}^{T}\mathbf{A}\mathbf{c}/2$. The relationship matrix **A** can be replaced by a genomic matrix **G**. Then it brings the question that whether the constraint should be based on pedigree (**A**) or genomic **(G)** relationship, when combined with breeding value estimation based on pedigree (EBV) or genomic information (GEBV).

The advantage of using GOCS in controlling in inbreeding was explained by Sonesson *et al.* (2012). When the selection was for GEBVs, the genomic inbreeding exceed expected pedigree inbreeding because the estimates based on pedigree information measure the inbreeding at a neutral locus not linked to any QTL, which does not exist in the genome. When the pedigree rate of inbreeding ΔF was controlled as expected at 0.5%, the genomic rate of inbreeding was ~0.7%. When the genomic relationship matrix was used to control genomic inbreeding at 0.5%, the pedigree inbreeding reduced to 0.4%. Thus using GOCS has a superior ability of restricting the rate of inbreeding compared to AOCS. In addition, Pryce *et al.* (2012) supported this theory with a fact that using a genomic relationship matrix (GRM) was the most effective in reducing average progeny inbreeding compared to using pedigree-based relationship matrix and shared runs of homozygosity (ROH).

According to Meuwissen *et al.* (2018), the main problem of whether to using GOCS or AOCS to control inbreeding depends on whether the inbreeding should be measured on neutral-linked loci or neutral-unlinked loci. Pedigree-based inbreeding coefficients denote the probability that two randomly chosen alleles from an individual are identical-by-descent (IBD). It assumes a truly neutral locus that is not affected by selection and not linked to a locus that is affected by selection. Thus AOCS limits the rate the rate of inbreeding at neutral unlinked loci, but not at neutral linked loci. However, in genomic selection schemes, the neutral-unlinked loci do not exist since the selection is for genome-wide dense marker data. Allele frequency changes at marker loci are substantially less systematic, which leads to more inbreeding at QTL than at marker loci. Thus GOCS can restrict genetic drift at marker loci and achieve genetic gain at QTL loci. Alternative genomic relationship matrices based on ROH is recommended since it

is a good indicator of true IBD status (de Cara *et al.* 2013). In addition, whole genome sequence (WGS) data may be a good information resource for GOCS, since it contains all the fitness, disease and loci of potential future interest.

In **chapter 2**, we firstly developed and tested the advanced OCS strategy based on pedigree information. Currently, the genomic information was available for some Angler sires. Thus in this project, for evaluating the long-term performance of different selection strategies (**chapter 4** & **chapter 5**), marker data was used for breeding value estimation and constructing kinship matrices instead of using traditional pedigree information. In **chapter 1**, we compared three methods of calculating the inbreeding coefficients: using pedigree information ($F_{PED}$), using G-matrix ($F_{GRM}$) and using runs of homozygosity ($F_{ROH}$). $F_{ROH}$ has a better performance of evaluating inbreeding since it captures both ancient and recent inbreeding. Thus in the following chapters (**chapter 4** & **chapter 5**) the kinship matrices of selection candidates were based on ROH instead of constructing G-Matrix. Although both pedigree and genomic information are involved in this project, for evaluating the efficiency of advanced OCS strategy, there is a lack of parallel comparison between using pedigree and genomic information resources.

de Cara *et al.* (2011) showed that for maintaining neutral genetic diversity, the most powerful strategy was to minimize the molecular coancestry computed on a SNP-by-SNP approach based on simulation studies. However, this strategy may lead to a decrease in fitness since maintaining genetic diversity also implies maintaining deleterious alleles. Using kinship based on ROH is more suitable for achieving a good balance between maintaining diversity and fitness. The only drawback is that phases of the genotype need to be estimated (Gómez-Romano *et al.* 2016).

### 6.4.2 Mating strategies for small populations

The selective breeding scheme includes two steps in practice: selection and mating. For the long-term impact of advanced OCS strategies, our simulation studies (**chapter 4** and **chapter 5**) used a very basic random mating system and the simulated generations are discrete. In the practical breeding scheme, choosing and implementing appropriate mating strategies after selection are worthwhile compared to random mating for two reasons. Firstly, choosing appropriate parent pairs may further increase genetic gain and/or reduce inbreeding (Caballero *et al.* 1996; Sonesson and Meuwissen 2000; Nirea *et al.* 2012). Secondly, any additional genetic

gain or reduction in inbreeding due to the implementation of suitable mating plans without additional costs or practical constraints can be treated as an additional benefit.

Wright (1921) proposed minimum-coancestry mating (MC mating), which select mating pairs by minimizing the average coancestry. When the parents are truncation-selected, minimum-coancestry mating generates less inbreeding compares to random mating because the onset of inbreeding level was delayed via minimizing the level of inbreeding in the next generation (Sonesson and Meuwissen 2000). The ancestral genetic contributions become more independent prior to stabilization that the genetic contribution of animals with higher Mendelian sampling terms are more likely to be higher (Woolliams *et al.* 2002). However, there is no guarantee the optimum mating can be found to minimize the rate of inbreeding because it only uses the degree of coancestry, which is not directly working with the contribution to increase independence. Henryon *et al.* (2009) proposed mating by minimizing the covariance of ancestral contributions (MCAC mating) which aims at minimizing the sum of absolute values of the covariances between ancestral genetic contributions. Based on the pedigree information, MCAC mating achieved as same genetic gain with MC mating with 4% to 8% reduction in inbreeding in the meantime. It was because MCAC mating increased the number of ancestors that made a long-term genetic contribution to the descendants of a population. Furthermore, Liu *et al.* (2017) reported that through simulation studies, minimum-coancestry and MCAC mating with genomic information realize a lower rate of inbreeding than with pedigree information without compromising rate of genetic gain ($\Delta G$). The reduction extent in the rate of inbreeding may be influenced by mating ratio, litter size and heritability of the trait. However, based on genomic information, the performance of MC and MCAC are similar in $\Delta F$ and $\Delta G$, even though the method of allocating mate pairs are different for both methods. That may be due to both methods not only controls the variance but also the covariance of genetic contributions.

So far all the above studies were based on the truncation selection scheme. Henryon *et al.* (2009) proposed that MCAC mating can be beneficially coupled with OCS strategy to achieve better performance. This combination increases the contribution of animals with the highest Mendelian sampling terms while having less influence on average coancestry, thus the genetic contributions become more independent. The benefit needs to be verified by further studies.

## 6.5 Concluding remarks

Genomic selection schemes should be implemented for local livestock breeds for the better prediction of both kinships among selection individuals and breeding values. Although the high prediction accuracy requires large reference population size, a joint population with close relationship may be applicable. For German Angler cattle, due to the relatively close relationship with Holstein cattle and red cattle breeds, it is promising to merge this cattle breed to construct a larger joint reference population for more accurate and reliable prediction. Due to the historical introgression, selection candidates from OCS procedure may contain a large amount of genetic material inherited from Holstein ancestors. Thus the advanced OCS strategy is necessary for balancing both production and conservation.

# Reference

Amador C., Toro M. Á., Fernández J., 2011 Removing exogenous information using pedigree data. Conserv. Genet. 12: 1565–1573.

Berg P., Nielsen J., Sørensen M. K., 2006 EVA: realized and predicted optimal genetic contributions. In: *Proceedings of the 8th world congress on genetics applied to livestock production:* 13–18 August 2006; Belo Horizonte.

Brøndum R. F., Rius-Vilarrasa E., Strandén I., Su G., Guldbrandtsen B., *et al.*, 2011 Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations. J. Dairy Sci. 94: 4700–4707.

Caballero A., Santiago E., Toro M. Á., 1996 Systems of mating to reduce inbreeding in selected populations. Anim. Sci. 62: 431–442.

Calus M. P. L., Mulder H. A., Bastiaansen J. W. M., 2011 Identification of Mendelian inconsistencies between SNP and pedigree information of sibs. Genet. Sel. Evol. 43: 34.

Calus M. P. L., Haas Y. de, Veerkamp R. F., 2013 Combining cow and bull reference populations to increase accuracy of genomic prediction and genome-wide association studies. J. Dairy Sci. 96: 6703–6715.

Daetwyler H. D., Capitan A., Pausch H., Stothard P., Binsbergen R. Van, *et al.*, 2014 Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat. Genet. 46: 858–865.

Dagnachew B. S., Meuwissen T. H. E., 2014 An iterative algorithm for optimum contribution selection in large scale breeding programs. In: *Proceedings of the 10th world congress on genetics applied to livestock production*: 17–22 August 2014; Vancouver.

Dagnachew B. S., Meuwissen T. H. E., 2016 A fast Newton–Raphson based iterative algorithm for large scale optimal contribution selection. Genet. Sel. Evol. 48: 48–70.

de Cara M. Á. R., Fernández J., Toro M. Á., Villanueva B., 2011 Using genome-wide information to minimize the loss of diversity in conservation programmes. J. Anim. Breed. Genet. 128: 456–464.

de Cara M. Á. R., Villanueva B., Toro M. Á., Fernández J., 2013 Using genomic tools to maintain diversity and fitness in conservation programmes. Mol. Ecol. 22: 6091–6099.

Doekes H. P., Veerkamp R. F., Bijma P., Hiemstra S. J., Windig J. J., 2018 Trends in genome-wide and region-specific genetic diversity in the Dutch-Flemish Holstein-Friesian breeding program from 1986 to 2015. Genet. Sel. Evol. 50: 15.

Erbe M., Hayes B. J., Matukumalli L. K. K., Goswami S., Bowman P. J., *et al.*, 2014 Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 97: 6622.

Ferenčaković M., Hamzić E., Gredler B., Solberg T. R., Klemetsdal G., *et al.*, 2013 Estimates of autozygosity derived from runs of homozygosity: Empirical evidence from selected cattle populations. J. Anim. Breed. Genet. 130: 286–293.

Gao H., Madsen P., Nielsen U. S., Aamand G. P., Su G., *et al.*, 2015 Including different groups of genotyped females for genomic prediction in a Nordic Jersey population. J. Dairy Sci. 98: 9051–9059.

Goddard M. E., Hayes B. J., 2009 Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10: 381–391.

Gómez-Romano F., Villanueva B., Sölkner J., Cara M. Á. R. de, Mészáros G., *et al.*, 2016 The use of coancestry based on shared segments for maintaining genetic diversity. J. Anim. Breed. Genet. 133: 357–365.

Hayes B. J., Bowman P. J., Chamberlain A. C., Verbyla K., Goddard M. E., 2009 Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet. Sel. Evol. 41: 51.

Henryon M., Sørensen A. C., Berg P., 2009 Mating animals by minimising the covariance between ancestral contributions generates less inbreeding without compromising genetic gain in breeding schemes with truncation selection. Animal 3: 1339–1346.

Hinrichs D., Calus M. P. L., Koning D. J. De, Bennewitz J., Meuwissen T. H. E., *et al.*, 2017 Biodiversity within and between European Red dairy breeds – conservation through utilization. In: *Proceedings of the 68$^{th}$ annual meeting of the European Association for Animal Production (EAAP)*: 28 Aug - 1 Sep 2017; Tallinn

Iheshiulor O. O. M., Woolliams J. A., Yu X., Wellmann R., Meuwissen T. H. E., 2016 Within- and across-breed genomic prediction using whole-genome sequence and single nucleotide polymorphism panels. Genet. Sel. Evol. 48: 15.

Liu Z., Seefried F. R., Reinhardt F., Rensing S., Thaller G., *et al.*, 2011 Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. Genet. Sel. Evol. 43: 19.

Liu H., Henryon M., Sørensen A. C., 2017 Mating strategies with genomic information reduce rates of inbreeding in animal breeding schemes without compromising genetic gain. Animal 11: 547–555.

Lund M. S., Ross S. P. W. de, Vries A. G. de, Druet T., Ducrocq V., *et al.*, 2011 A common reference population from four European Holstein populations increases reliability of genomic predictions. Genet. Sel. Evol. 43: 43.

Lund M. S., Su G., Janss L., Guldbrandtsen B., Brøndum R. F., 2014 Invited review: Genomic evaluation of cattle in a multi-breed context. Livest. Sci. 166: 101–110.

Lund M. S., Berg I. Van den, Ma P., Brøndum R. F., Su G., 2016 Review: How to improve genomic predictions in small dairy cattle populations. Animal 10: 1042–1049.

Mészáros G., Boison S. A., Pérez O'Brien A. M., Ferenčaković M., Curik I., *et al.*, 2015 Genomic analysis for managing small and endangered populations: A case study in Tyrol Grey cattle. Front. Genet. 6: 1–12.

Meuwissen T. H. E., 1997 Maximizing the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75: 2575–2583.

Meuwissen T. H. E., Hayes B. J., Goddard M. E., 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829.

Meuwissen T. H. E., 2002 GENCONT: an operational tool for controlling inbreeding in selection and conservation schemes. In: *Proceedings of the 7$^{th}$ world congress on genetics applied to livestock production*: 19–23 August 2002; Montpellier.

Meuwissen T. H. E., Sonesson A. K., Woolliams J. A., 2018 Genomic management of inbreeding in breeding schemes. In: *Proceedings of the 11th world congress on genetics applied to livestock production*: 12–16 Feb 2018; Auckland.

Mullin T. J., Belotti P., 2016 Using branch-and-bound algorithms to optimize selection of a fixed-size breeding population under a relatedness constraint. Tree Genet. Genomes 12: 1–12.

Mullin T. J., 2017 *OPSEL 2.0: a computer program for optimal selection in tree breeding*.

Nirea K. G., Sonesson A. K., Woolliams J. A., Meuwissen T. H. E., 2012 Effect of non-random mating on genomic and BLUP selection schemes. Genet. Sel. Evol. 44: 11.

Olson K. M., VanRaden P. M., Tooker M. E., 2012 Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. J. Dairy Sci. 95: 5378–5383.

Pfaff B., 2014 The R package cccp: Design for solving cone constrained convex programs. R Financ.: http://www.pfaffikus.de/files/conf/rif/rif2014.pdf.

Pong-Wong R., Woolliams J. A., 2007 Optimisation of contribution of candidate parents to maximise genetic gain and restricting inbreeding using semidefinite programming. Genet. Sel. Evol. 39: 3–25.

Pryce J. E., Hayes B. J., Goddard M. E., 2012 Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. J. Dairy Sci. 95: 377–88.

Raymond B., Bouwman A. C., Schrooten C., Houwing-Duistermaat J., Veerkamp R. F., 2018 Utility of whole-genome sequence data for across-breed genomic prediction. Genet. Sel. Evol. 50: 27.

Sonesson A. K., Meuwissen T. H. E., 2000 Mating schemes for optimum contribution selection with constrained rates of inbreeding. Genet. Sel. Evol. 32: 231–248.

Sonesson A. K., Woolliams J. A., Meuwissen T. H. E., 2012 Genomic selection requires genomic control of inbreeding. Genet. Sel. Evol. 44: 27.

Stålhammar H., 2018 Improved genetic progress after introduction of genomic selection – a red dairy cattle example. In: *Proceedings of the 11th world congress on genetics applied to livestock production*: 12–16 Feb 2018; Auckland.

Thomasen J. R., Egger-Danner C., Willam A., Guldbrandtsen B., Lund M. S., *et al.*, 2014a Genomic selection strategies in a small dairy cattle population evaluated for genetic gain and profit. J. Dairy Sci. 97: 458–470.

Thomasen J. R., Sørensen A. C., Lund M. S., Guldbrandtsen B., 2014b Adding cows to the reference population makes a small dairy population competitive. J. Dairy Sci. 97: 5822–5832.

VanRaden P. M., 2008 Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414–23.

Wellmann R., 2018 Optimum contribution selection and mate allocation for breeding: the R package optiSel. submitted.

Woolliams J. A., Pong-Wong R., Villanueva B., 2002 Strategic optimisation of short- and long-term gain and inbreeding in MAS and non-MAS schemes. In: *Proceedings of the 7th world congress on genetics applied to livestock production*: 19–23 August 2002; Montpellier.

Woolliams J. A., Berg P., Dagnachew B. S., Meuwissen T. H. E., 2015 Genetic contributions and their optimization. J. Anim. Breed. Genet. 132: 89–99.

Wright S., 1921 Systems of Mating. I. the Biometric Relations between Parent and Offspring. Genetics 6: 111–123.

Zhou L., Heringstad B., Su G., Guldbrandtsen B., Meuwissen T. H. E., *et al.*, 2014a Genomic predictions based on a joint reference population for the Nordic Red cattle breeds. J. Dairy Sci. 97: 4485–4496.

Zhou L., Lund M. S., Wang Y., Su G., 2014b Genomic predictions across Nordic Holstein and Nordic Red using the genomic best linear unbiased prediction model with different genomic relationship matrices. J. Anim. Breed. Genet. 131: 249–257.

# LIST OF PUBLICATIONS

## Peer-Reviewed Journal Publications

**Yu Wang**, Jörn Bennewitz, Robin Wellmann*: Recovering the original genetic background of local breeds with historical introgression using advanced optimum contribution selection strategies.* (Submitted for publication)

**Yu Wang,** Patrick Stratz, Dierck Segelke, Robin Wellmann, Dirk Hinrichs, Jörn Bennewitz: *Genome-wide estimates of inbreeding and characteristics of runs of homozygosity in German Angler cattle*. (In preparation)

**Yu Wang**, Dierck Segelke, Reiner Emmerling, Jörn Bennewitz, Robin Wellmann: *Long-term impact of optimum contribution selection strategies on local livestock breeds with historical introgression using the example of German Angler cattle*. G3: GENES, GENOMES, GENETICS December 1, 2017 vol. 7 no. 12 4009-4018; doi: 10.1534/g3.117.300272

**Yu Wang**, Jörn Bennewitz, Robin Wellmann: *Novel optimum contribution selection methods accounting for conflicting objectives in breeding programs for livestock breeds with historical migration*. Genetics Selection Evolution 05/2017; 49:45. doi:10.1186/s12711-017-0320-7

## Conference Contributions

Robin Wellmann, **Yu Wang**, Jörn Bennewitz: *Removal of haplotype segments originating from foreign breeds using optimum contribution selection.* Book of Abstracts of the 69th Annual Meeting of the European Federation of Animal Science; Dubrovnik, Croatia; 08/2018

**Yu Wang**, Jörn Bennewitz, Robin Wellmann: *Advanced optimum contribution selection strategies in the conservation of local livestock breeds with historical introgression*. In Proc of 11th WCGALP, Auckland, New Zealand; 02/2018

**Yu Wang**, Jörn Bennewitz, Robin Wellmann: *Long term impact of optimum contribution selection on breeds with historical introgression.* Book of Abstracts of the 68th Annual Meeting of the European Federation of Animal Science (pp. 81-81); Tallinn, Estonia; 08/2017

**Yu Wang**, Jörn Bennewitz, Robin Wellmann: *Comparison of optimum contribution selection methods at the example of two local cattle breeds*. Vortragstagung der DGfZ und GfT, Hannover, Germany; 09/2016

Marleen Visker, **Yu Wang**, Mathijs van Pelt, Henk Bovenhuis: *Not sex but birth weight of the calf affects its mother's milk production*. 12th International Symposium on Milk Genomics and Human Health, Sydney, Australia; 10/2015

Marleen Visker, **Yu Wang**, Mathijs van Pelt, Henk Bovenhuis: *Milk production of the mother is associated with the birth weight rather than the sex of the calf*. Book of Abstracts of the 66th Annual Meeting of the European Federation of Animal Science (pp. 522-522), Warsaw, Poland; 08/2015

# ACKNOWLEDGEMENT

At the end of this thesis, I would like to take some time to thank all the people without whom this project would never have been possible. Although it is just my name on the cover, many people have contributed to this thesis in their own particular way and for that, I want to thank you all.

First of all, I am extremely grateful to my supervisors, Professor Jörn Bennewitz and Dr. Robin Wellmann. Thank you for offering this invaluable space for me to participate in this project and develop myself as a researcher in the best possible way. Thanks for the guidance and all the discussions and brainstorming sessions. The deep insights helped me at various stages of my research. Furthermore, I am grateful for your understanding and continuous encouragement during the times when I was really unconfident and frustrated due to the challenges I was facing. Your steadfast support and intellectual guidance—from start to finish— have been very helpful and I feel incredibly privileged to have you as my supervisors. I have to thank the German Research Foundation for the financial support of this project and Hanna for assisting me with the German summary part in the thesis. In addition, I also want to thank my examination committee. Thank you for investing time and providing interesting and valuable feedback for this thesis.

And then, of course, my lovely Hohenheim colleagues. Thanks for the fantastic team I have been fortunate to work with. From day one I arrived here, you guys are so angel-hearted and helped me with everything especially the complicated paperwork that I never need to worry about how to deal with it. The group has been a source of friendships as well as good advice and collaboration. And of course, I will never forget the time with cutie dogs around every day. I know it is a rather short journey for me to work in this awesome group, the memory is so remarkable that I will keep it in my heart.

I always feel grateful that I am very lucky to have all my supervisors, who not only impart knowledge to me but also offer me professional advice in my research career. I have to thank Prof. Yunhai Zhang, who persuades me to insist on my dream since my bachelor time. Without him, I would never imagine there is one day I could come to study in Europe. Very special thanks to Prof. Henk Bovenhuis and Dr. Marleen Visker for supervising my first master thesis

in Wageningen, and also to Prof. Johann Sölkner and Dr. Gábor Mészáros for supervising my second master thesis in BOKU, Vienna. This special experience of studying in the EMABG program shaped my research interest and encouraged me of pursuing a Ph.D. degree in animal breeding and genetics. I would also like to take this opportunity to thank Dr. Solomon Antwi Boison, for his endless encouragement, support, patient guidance and illuminative ideas at the beginning of my research career. Last but not least, a heartfelt thanks to my friends who made my European experience something special, in particular, Marta, Rajesh, Sanne, Harmen, Adrien, Karina, Tesfaye, Christina, Stefana, Sterre, Marjolein, Devan, Stacey (along with all the lovely friends in New City Wien), Severin, Shijie, Hang, Zoe, Ally, Luping and Cheng. I also want to particularly thank Mandy, Lifei and Shizhi, for their precious advice in both of work and life for the past five years. I am very blessed to have you all.

I am also indebted to my besties Alice and Jingyao, not only for their endless support and encouragement for the past years, but also for being there to listen when I needed an ear, regardless of the seven hours' time difference between each other. I will never forget those dispirited and stressful nights that I couldn't sleep, but still, having your comforting words around warms me a lot.

Of course, words cannot express the feelings I have for my parents for their constant unconditional support - both emotionally and financially. I feel lucky for being your daughter.

Thank you all:)

Yu

# CURRICULUM VITAE

## PERSONAL DETAILS

| | |
|---|---|
| Name | Yu Wang |
| Date of Birth | 18.02.1990 |
| Place of Birth | Nei Mongol, China |
| Nationality | Chinese |

## EDUCATION

| | |
|---|---|
| Nov 2015 – to date | **Ph.D. Department of Animal Genetics and Breeding**<br>University of Hohenheim<br>Stuttgart, Germany |
| Aug 2013 – Jul 2015 | **M.Sc. Animal Sciences (Animal Breeding and Genetics)**<br>University of Natural Resources and Life Science, Vienna<br>Vienna, Austria |
| Aug 2013 – Jul 2015 | **M.Sc. Animal Sciences (Animal Breeding and Genetics)**<br>Wageningen University & Research<br>Wageningen, the Netherlands |
| Sep 2007 – Jul 2011 | **B.Sc. Applied Biological Science**<br>Anhui Agricultural University<br>Hefei, China |

## COURSES

| | |
|---|---|
| Aug 2016 | **Quantitative Genetics in Animal Breeding**<br>University of Helsinki<br>Mikkeli, Finland |
| Feb 2014 | **Getting Started in ASReml**<br>Wageningen University & Research<br>Wageningen, the Netherlands |